PEPTIDE SCAFFOLD CONJUGATES AS PROTEIN MIMICS FOR PHOSPHATE RECOGNITION

YARA RUIZ GARCÍA

Promoter: Prof. Dr. Annemieke Madder

Dissertation submitted in fulfillment of the requirements for

the degree of Doctor of Philosophy (PhD)

DOCTOR OF SCIENCE: CHEMISTRY

2014

APPRECIATION AND ACKNOWLEDGEMENTS

As I finalize my thesis as a last step towards becoming a doctor in chemistry, I feel as if a goal and dream that started very early ever since I was a child will finally be fulfilled. My interest for nature and science in general was the keystone behind my passion for chemistry and research. Scientists play a fundamental role in the development of new therapies for the treatment of human diseases. The fact that scientists can have an ever-lasting impact on people's well-being encouraged me to focus my research in this area. My humble contribution to this vast ocean of knowledge is reflected in this document. This achievement, however, would not have been possible without the contribution of several people. These people have accompanied me on this long and difficult yet exciting and rewarding journey, being by my side through bad and good. Therefore, they should be mentioned in this thesis.

First of all, I would like to thank my promoter, Prof. Dr. Annemieke Madder, for giving me the opportunity to start a PhD in her group. Without her unwavering support during the failures I had with the click reaction and her enthusiasm for science which always encouraged me to go on despite the challenges throughout my projects this PhD could not have been completed. Most importantly, if I can communicate well in English and present before you my work today with confidence, it is only because she emphasized time and again the importance of these skills.

I would also like to acknowledge those who have been close to me in the lab during my PhD work. Abhishek, for your willingness to learn during your master thesis and later for being a colleague and sharing projects and ideas. I will always be grateful for this. Your help in Lincoln and your support while I was writing the thesis was invaluable. Eva, you were there during all the crazy shifting I did when I moved to the UK and then back to Belgium, offering me your place to crash. I would like to mention all my master students Michiel, Eveline, Vincent and Cecilia, for all the good work you have done. You are all unique and special. I am glad to say that I got along with all of you very well and will always be proud of your achievements. I am also happy that you are doing well in your respective careers. It has been a pleasure to have several useful conversations about chemistry with An, Duchan, Matthias, Dorien and Lidia. Thanks to Kurt, Margarida, Vicky, Nathalie, Lars, Eri, Willem, Marieke, Bram, Maxime, Ellen, Smita and Emily for the day-to-day work in the lab. Last but not least, I would like to thank Jos for always being there and explaining practical things to us in a friendly manner and fixing our equipment, whenever technical problems occurred.

As for the Organic and Macromolecular Chemistry Department, I would like to extend my thanks to Prof. Jose Martins, Tim, Dieter and Davy for the useful discussions related to NMR and concentration determination. Thanks to Jan for the countless analyses performed during these 3 years and 3 months of my

PhD. His input and advices about analysis were always helpful. Thanks also to Veerle for her kind assistance when ordering compounds and later for her help with administrative tasks.

Several collaborations were established by Annemieke and myself during the course of this PhD thesis, many of which resulted in excellent projects and ideas. The people involved in these projects deserve to be mentioned for their contribution. Prof. Edvard Smith and Vladimir Pabón Martínez from Karolinska Institute, Sweden for the EMSA studies. Prof. Tomas Kraus and Jan Zelenka from the Academy of Sciences of the Czech Republic for the synthesis of the cyclodextrin derivatives. Prof. Bruno De Geest, Dorien and Benoit Louage from Ghent University for testing our compounds *in cellulo*. Prof. Gilles Bruylants and Simon Croegaert from Université Libre de Bruxelles are also acknowledged for their help during the ITC measurements.

In addition, I would like to allude to the European Commission for their financial support. This thesis has been funded by a Marie Curie Initial Training Network. As a Marie Curie fellow of the PhosChemRec Network, I could work together with other European Institutions and discuss projects with PhD students from other countries. A big thanks to Marta, Alice, Zeyed, Plamena, Martina, Charlotte, Emmanuel, Chris, Sylvain, Cristina, Ute, Daniela, Luigi, and Pierre-Yves for the time we spent together in scientific meetings, discussing amongst us about chemistry and sharing personal experiences.

Due to financial constraints I could not be funded in my final year as a PhD. However, this was a blessing in disguise as it gave me the opportunity to start a new research project as a research assistant in the University of Lincoln (UK). Prof. Ishwar Singh and Edward Taylor are accredited for the supervision of the project and Charlotte Bolt for the useful discussions about chemical biology. Hugs to my friends Gio, Isa and Martin for our spare time together and for encouraging me to write my thesis after working hours and in weekends during my 6 month stay in Lincoln.

As members of the jury, I would like to thank Dr. Ishwar Singh, Prof. Johan Winne and Prof. Annemieke Madder for carefully reading and correcting this thesis. Prof. Roger Strömberg, Prof. Bruno De Geest, Prof, Gilles Bruylants and Prof. Richard Hoogenboom as chairman are also thanked for taking their valuable time and agreeing to be part of my jury.

Then, I would like to mention my non-lab friends in Ghent who have been an integral part of PhD life. Thanks to María, Julián, Zulma, Víctor, Amparo, Conchi, Nuria, Joana, Trompi, Miguelito, Paco, Julia… for all the good moments together. A special mention to my friends I met during my dancing lessons and social events. Our nice time together has given me countless memories to cherish. My final paragraphs will be in Spanish in fond memories of my family and home town.

Me gustaría agradecer también a mis amigos de Palencia, a Kike, Jorge, Alicia, Pirri, Ana Luz, Rafi, Sevi, Burgos, Cris, Rafa… por apoyarme cuando tomé la decisión de hacer el doctorado en el extranjero. A pesar de la distancia habéis estado a mi lado y cada vez que he vuelto a Palencia me he sentido como si nunca me hubiera marchado. Gracias por guardar nuestra amistad. También me gustaría agradecer a mi colegio, La Salle, por la educación que recibí, tanto académica como personal. Gracias a Juanan, Nano, Javi, Ana y

Marta por todo el tiempo que han dedicado a los grupos cristianos del colegio. Me gustaría hacer una mención especial para el Hno. Tomás González, mi primer profesor de química. Gracias por sus clases, su ayuda a los alumnos en horas extra-escolares y su pasión por la ciencia, que fue lo que me impulsó a seguir con mi idea de estudiar química.

Por último pero no por ello menos importante, me gustaría expresar la gratitud que siento hacia mi familia, a mi madre y a mi padre por enseñarme a ser perseverante y constante, a exigirme lo máximo de mi misma, a ser ambiciosa y conseguir mis objetivos sin tener que pisar a nadie para ello. Gracias por enseñarme el respeto a los demás, a saber ayudar a los que lo necesitan y a nunca menospreciar a nadie. Todos estos valores, son desde mi punto de vista fundamentales para poder dar lo mejor de uno mismo y trabajar en grupo. Gracias también por haberme apoyado en los momentos difíciles. Me gustaría dar las gracias a mi hermano Javi, porque desde que llegué a Bélgica hemos pasado por muchas situaciones, desde frotar paredes hasta irnos de boda y ampliar la familia, estando siempre juntos. Gracias Elaine por tu apoyo y optimismo en todo lo que ha pasado estos últimos años. A mi sobrina Nora porque desde lo simple me has sabido transmitir calma y felicidad. Me gustaría mencionar a Liam, mi sobrino, para que sepas que te estamos esperando. Y por último, gracias a mis tíos, a mis primos y a mi abuela, por haber confiado en mí.

Yara Ruiz García

Ghent, 19th December 2014

STRUCTURES AND ABBREVIATIONS

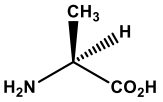
AA	amino acid
Aba	4-acetamido benzoic acid
AcIm	1-acetylimidazole
Ac_2O	acetic acid anhydride
AcOH	acetic acid
Alloc	allyloxycarbonyl (protecting group)
Boc	tert-butoxycarbonyl (protecting group)
BSA	bovine serum albumin
b-HLH-ZIP	basic helix-loop-helix zipper
bZIP	basic region leucine zipper
CD	cyclodextrin
CuAAC	copper azide-alkyne cycloaddition
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DHB	2,3-dihydroxybenzoic acid
DIPEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	N, N-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (also as hydrochloride)
EDT	ethanedithiol
EMSA	electrophoretic mobility shift assay

ESI	Electrospray ionization
EtOAc	ethylacetate
EM	exact mass
Fmoc	9-fluorenylmethoxycarbonyl
Gaba	γ-aminobutylbenzoic acid
HATU	N-[(dimethylamino-1H-1,2,3-triazolo[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide
HBTU	N-[(1H-benzotrioazole-1-yl)(dimethylamino) methylene]-N-methylmethanaminium hexafluorophosphate N-oxide
HCA	α-cyano-4-hydroxycinnaminic acid
hν	UV-light or photocleavable linker
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
HPA	hydroxypicolinic acid
HPLC	High-performance liquid chromatography
HRMS	high resolution MS
ICL	interstand cross-link
LC	Liquid Chromatography
MALDI	matrix assisted laser desorption ionisation
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
MeCN	acetonitrile
MeOH	methanol
mRNA	messenger RNA
MS	Mass Spectrometry
MW	molecular weight or microwave
mQ milliQ,	deionized H_2O
m/z	mass to charge ratio

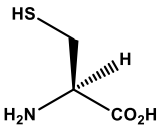
NMP	N-methyl-2-pyrrolidine
NMR	nuclear magnetic resonance
NP	NovaPEG
Pbf	2,2,4,6,7-Pentamethyl-dihydrobenzofurane-5-sulfonyl (protecting group)
PDB	protein data bank
PEG	polyethyleneglycol
ppm	parts per million
PS	phosphatidylserine
PS	polystyrene
PSS	phosphorothiosulfide
PyBOP	(Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
Rf	ratio to front
rt	room temperature
r.t.	retention time
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RP	reversed phase
SPPS	Solid Phase Peptide Synthesis
ssDNA	single stranded DNA
TBE	TrisBoratEDTA (buffer) or tris(hydroxymethyl aminomethane, B(OH)3, ethylenediamine tetraacetic acid
TCEP	tris(2-carboxyethyl)phosphine
tBu	tert-butyl (protecting group)
TF	transcription factor
TFA	trifluoro acetic acid
TG	TentaGEL
TIS	triisopropylsilane

TNBS	trinitrobenzenesulfonic acid
Tris	Trizma hydrochloride buffer solution
tRNA	transfer RNA
Tris	tris(hydroxymethyl)aminomethane)
Trt	trityl or triphenylmethyl (protecting group)
UV	Ultraviolet

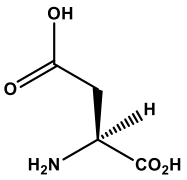
List of amino acids



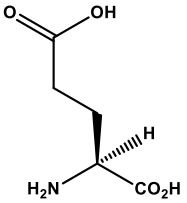
Alanine, Ala, A



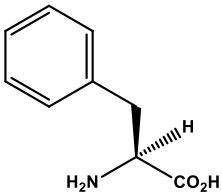
Cysteine, Cys, C



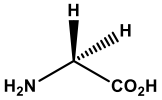
Aspartic Acid, Asp, D



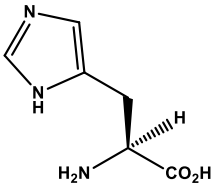
Glutamic Acid, Glu, E



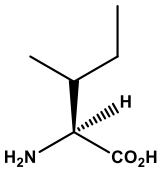
Phenylalanine, Phe, F



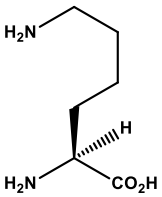
Glycine, Gly, G



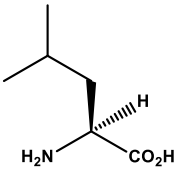
Histidine, His, H



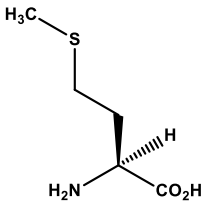
Isoleucine, Ile, I



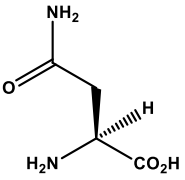
Lysine, Lys, K



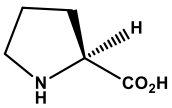
Leucine, Leu, L



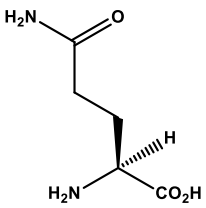
Methionine, Met, M



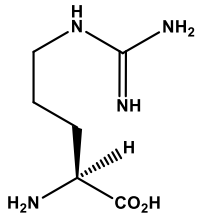
Asparagine, Asn, N



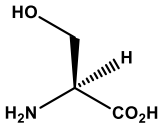
Proline, Pro, P



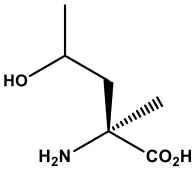
Glutamine, Gln, Q



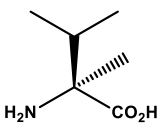
Arginine, Arg, R



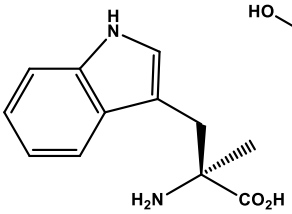
Serine, Ser, S



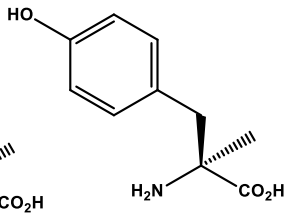
Threonine, Thr, T



Valine, Val, V



Tryptophan, Trp, W



Tyrosine, Tyr, T

PRELUDE

The first step towards the discovery of the double helical structure of DNA by Watson and Crick has its origins in the end of S.XIX, when Friedrich Miescher, interested in the substructure of the cell, studied the identification of proteins in the nucleus. He was able to isolate nuclei from human leukocytes. Furthermore, by means of chemistry and biology, Miescher could analyse the purified nuclei with the chemical tools available then: digestion by proteases, elemental analysis and solubility tests, obtaining what he defined as:

In the experiment with the weakly alkaline fluids, I obtained, by neutralization of the solutions, precipitates which were insoluble in water, acetic acid, very dilute hydrochloric acid, or sodium chloride solutions; consequently, they could not belong to any of the known albuminoid substances. Where did this substance come from?[1]

The elemental composition of the new substance, which he named nuclein, included carbon, oxygen, hydrogen and nitrogen, elements commonly found in organic compounds, but also phosphorus. It was not until 1889 when Richard Altmann, one of the students of Miescher, defined nuclein as nucleic acid. The discovery of nucleic acids using chemical methods to unravel the mysteries of the cell can be considered as the origin of chemical biology in the contemporary era. Despite several outstanding findings related to the new substance, including the discovery of their components (phosphate, sugar and the four bases), the way they are connected through the phosphate group, forming a polymeric chain of nucleotide subunits, the isolation of DNA and RNA and the X-Ray studies of chromosomes and DNA, it was not until 1953 when Francis Crick and James Watson reported the first accurate model of DNA's molecular structure. They received the Nobel Prize in Physiology or Medicine in 1962 together with Maurice Wilkins *"for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"*.

But the history of DNA did not end with Watson and Crick resolving its three dimensional structure. The story was far from over. In the late 50s, it was known that DNA contained the hereditary information of the cell. Francis Crick reasoned that the sequence of nucleotides in the molecule must function as a code which directs the synthesis of proteins. By resolving the puzzle on how DNA leads to the synthesis of proteins that are made in the cytoplasm, he established the Central Dogma of molecular biology: The information contained in the DNA is transcribed to RNA, which serves as an information carrier to the cytoplasm, where RNA is translated into proteins. In 1968, Nirenberg, Khorana and Holley received the Nobel Prize in Physiology or Medicine for deciphering the genetic code and describing how it operates in protein synthesis. During the second half of last century, many contributions to the area of genetic information and its relation to cell function and structure were developed, but it was not until 2003 when the human genome

was completed, during what is known as the world largest biological research project, the Human Genome Project (HGP). Lead by universities and research centers in the United States, the United Kingdom, Japan, France, Germany and China, they sequenced and mapped all the genes, together known as genome, of members of the Homo sapiens species.

All the information revealed during more than 100 years can be translated today into new paths to investigate normal human biology, the aberrant processes and elaborate networks of gene, behaviour and environment that lead to human disease[2,3]. However, despite the revelation on how the genetic information is stored in the cell and translated into proteins, there is still a gap in the way this overwhelmingly large amount of data can be used to understand protein structure, function, interactions and cell localization, as a vast number of proteins are still uncharacterized. On the other hand, protein function can be determined by the cooperation between different disciplines. Traditionally, biology had predominance over other areas of science when concerning the cellular environment. Due to a large number of advancements in life sciences in recent years, interest in the nature of the molecules that constitute the essential components of life and their interactions has opened an emerging area for the understanding of the molecular basis of life. Biological processes can be explained by multiple related chemical reactions. Therefore, the convergence of biology and chemistry can result in potent developments towards the study of interactions between biomolecules and methodology to synthesize compounds that can mimic their functions. The current advances in synthetic and analytical chemistry allow the obtention and characterization of large sized molecules possessing specifically designed properties. These supramolecular entities can facilitate the comprehension of biological systems.

The areas of biology and chemistry are thus merging into one discipline, chemical biology, in which chemistry provides answers to questions otherwise not possible on the basis of pure biology. The challenge of exporting the well-known chemical reactions and mechanisms to biomolecules enables the development of new methodologies for bioorganic synthesis. Moreover, chemistry affords well-defined structural compounds for the development of novel biological analogs and the study of biological functions on a molecular scale. In addition, covalent modifications of materials such as peptides broaden the understanding of protein function and can even provide the biologist with new possibilities[4]. Unnatural modifications of proteins are not achievable by the ribosomal machinery in an efficient way. Thus, there is a need for compounds containing non-peptidic structural elements that are capable of mimicking or antagonizing the biological action of a natural parent peptide. For example, the non-peptidic-part can simplify the overall protein structure by substituting sectors of the protein not needed specifically for the biological function. Therefore, they are clustered into peptidomimetic compounds[1], and can be synthesized by any of the many chemical conjugations.

[1] For an exact description, a peptidomimetic is a compound that is designed to mimic the biological activity of a peptide, but with the desired structural modifications for the improvement of its use as a drug. Their structure resembles peptides but with the incorporation of artificial modifications that enable the obtention of the required properties.

In the current thesis, we aim to mimic biological macromolecules involved in key functions within the cell for both inhibiting and modulating their activity. In view of the expertise of the OBCR group in organic chemistry, peptide synthesis and bioconjugation strategies, we here present the combination of these tools for the successful synthesis of models of the bZIP and bHLH transcription factor families. These models are based on a peptidic part and a scaffold which allows the peptide to adopt the correct position for its function, and therefore are classified as peptidomimetic compounds.

INTRODUCTION

PHOSPHATE RECOGNITION – PAST, PRESENT & FUTURE

The complete role that phosphate anions play in the environment and biology is still very much a mystery despite countless efforts towards studying the selective binding of phosphate and phosphorylated molecules for more than 100 years. One of the first attempts for the recognition of phosphorus in biological materials was carried out by Taylor and Miller in 1914 [5]. Their strategy consisted of a method based on a colored molybdenum (IV) phosphate complex. Further improvements towards less time consuming and non-toxic practices focused on enzymes as biological elements that had affinity for the phosphate anion containing molecules. Unfortunately, enzymes suffer from poor stability and demanding production costs. As an alternative approach and in order to overcome these drawbacks, chemists proposed synthetic receptors for the extraction, transport and detection of phosphorylated molecules and phosphate anions. A large variety of artificial receptors with successful affinity for phosphates have been documented. Despite these merits, the sensitive and selective binding of phosphorylated molecules remains a challenging area of research in supramolecular chemistry.

Phosphate anions are present in molecules with diverse biological functions. Together with sugars, they constitute the backbone of DNA in which genetic information is stored. Phosphorylation of proteins occur during the posttranslational modification of the protein surface and is involved in switching the enzymatic activity for the generation of signaling transduction pathways. Phosphate anions are also present at the cellular membrane as phospholipids. The randomization of their distribution along the inner and the outer part of the membrane is related to the activation of hemostasis and thrombosis. In addition, cell death and clearance by phagocytosis, a process where apoptotic cells are removed from the bloodstream by macrophages that specifically recognize phosphatidylserine on the cell surface is also a process phosphate anions are involved in.

Through the course of this thesis, we aim to unravel some of the difficulties that chemists and biologists are currently facing regarding phosphate recognition. In chapter 1, we approach small phosphates, particularly phosphatidylserine, for the early detection of apoptosis. The following chapters will focus on DNA as target for the development of synthetic transcription factor models and artificial nucleases based on peptide conjugates. Research is a continuous process and every question answered only raises two more. It is this part which makes it challenging, exciting and worth doing. We hope our results will facilitate future efforts in this area and enable the development of receptors for phosphate containing molecules for broadening scientific knowledge and more specifically for biological applications in vivo.

TABLE OF CONTENTS

II. EXPERIMENTAL SECTION

PART I:

DESCRIPTIVE SECTION

CHAPTER 1

DETECTION OF PHOSPHATIDYLSERINE

The presence of phosphates in many molecules that constitute biological systems make them ideal candidates as targets to study and modulate biological functions. Therefore, receptors which could have potential for phosphate recognition have been thoroughly studied. Consequently, a broad range of recognition units have been developed for their interaction with small phosphate-containing molecules such as phospholipids as well as for macromolecules like DNA. During the course of this thesis an overview of previous synthetic receptors is presented, together with our contribution to this ever-growing area of research. In this chapter, we introduce our first approach for phosphate recognition by the design of synthetic receptors for phosphatidylserine for the early detection of apoptosis.

1.1 Phosphatidylserine

1.1.1 Structure and significance of phosphatidylserine

Phosphatidylserine (PS) is a glycerophospholipid present in the cellular membrane of all eukaryotic organisms. The structure of phosphatidylserine consists of a glycerol backbone esterified on the sn-1 and sn-2 carbons of the glycerol moiety, a phosphate group on sn-3 and two fatty acyl chains of variable length and saturation (figure 1.1). In comparison with other phospholipids present in the cell membrane, PS possesses a serine attached to the phosphate group and is the major anionic phospholipid in the plasma membrane. Due to the net negative charge of the hydrophilic head group, the combination of several PS molecules provides a net negatively charged layer which is exposed to the cytosol of the cells.

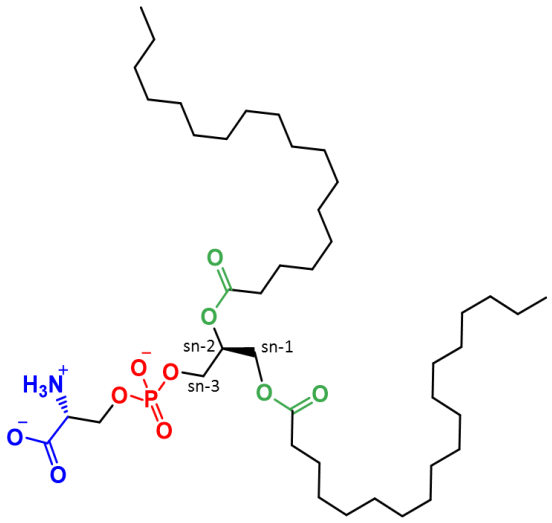


Figure 1.1: Structure of phosphatidylserine with the stereospecific numbering (sn) of the glyceric part.

The main roles of PS involve signaling in both an intra and an intercellular environment. The appearance of PS on the cell surface is one of the signals of the early stages of cell apoptosis[6] and blood clotting. On the other hand, the presence of PS at the intracellular surface of the membrane is essential for the localization and activation of proteins such as ubiquitin-protein ligase, protein kinases and phospholipases amongst others.

1.1.2 Localization of PS in the cellular membrane

The non-random distribution of lipids in the bilayer of the membrane is a common feature in eukaryotic cells and is a consequence of the role of particular enzymes. Due to the biophysical properties of lipids, they can spontaneously "flip" their polar headgroups through the hydrophobic interior of the membrane. In addition to this, the occurrence of active lipid translocation is carried out by transporters (enzymes) across the bilayer. These energy-independent and dependent processes are responsible for some of the biological functions in the cell, such as signal transduction, membrane fusion, blood coagulation, and cell apoptosis. Thus, membrane phospholipids have been targeted for the development of synthetic translocases as pharmaceuticals or as chemical tools for biological membrane research[7] and for the detection of early-stage apoptosis.

The enzymes responsible for the transport of phospholipids through the bilayer are called translocases. They can be divided in three groups: scramblases, which transport phospholipids in a bidirectional manner through the membrane. Flippases which are energy-dependent transporters and move phospholipids toward the inner surface of the membrane. Finally, floppases which are similar to flippases but move the phospholipids away from the inner surface of the membrane. The best known member of the flippase family is the aminophospholipid flippase which transports PS and to a lesser extent phosphatidylethanolamine (PE) from the outer to the inner monolayer[8] (figure 1.2).

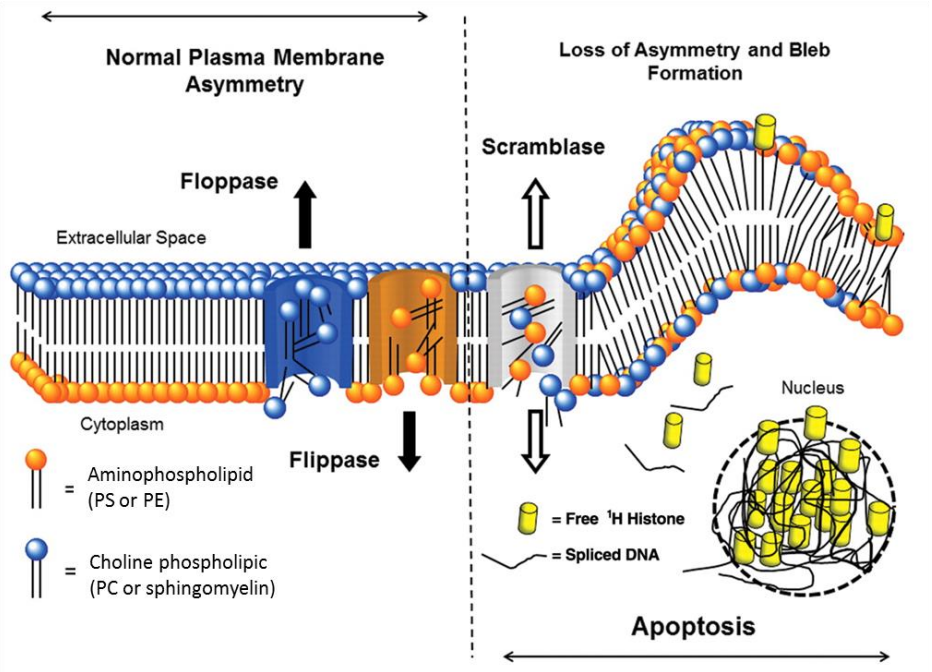


Figure 1.2[1]. Changes in cell membrane configuration with apoptosis. Redistribution of phospholipids assists in formation of membrane blebs and ultimately apoptotic bodies. Also of note is release of ^{1}H histone protein as DNA is spliced by caspase3. Free ^{1}Histones are translocated to cell surface by as of yet undescribed mechanisms[9].

In addition, due to the biological influence that PS has on apoptosis and blood clotting, there is a major ongoing effort among researchers to target PS for the detection of apoptosis as cell death control. Apoptosis is accepted as one of the key readouts for the early prediction of the efficiency of cancer therapies like chemotherapy, radiation or anti-angiogenic drugs. Therefore, strategies to detect apoptosis are of great interest for disease diagnosis and assessment of treatment responses. Thus, non-invasive imaging technologies can be developed based on the detection of PS[10–12].

In this chapter we present our first attempts towards the synthesis of artificial receptors for phosphatidylserine as probes for the detection of apoptosis.

1.2 Apoptosis

The term apoptosis is a Greek word that describes natural cell death. Over the years, due to the developments in cellular biology, it has been adjusted to programmed cell death. Apoptosis occurs in a cell when its functions are no longer needed. In addition, it helps to regulate the number of cells in the organism. In contrast to necrosis, where cell death takes place as the result of cellular injury, apoptosis is regulated by the organism and its dysregulation can lead to dramatic consequences such as the development and progression of cancer. The ability of cancer cells to avoid apoptosis and proliferate is one of the trademarks of cancer and therefore, also a main target of cancer treatment.

[1] Adapted from F. G. Blankenberg and J. F. Norfray, *AJR. Am. J. Roentgenol.*, 2011, **197**, 308–17.

Usually, PS resides in the inner part of the cell membrane. During apoptosis, the control of the asymmetry of the cellular membrane gets disrupted due to the transport of PS from the inner to the outer surface. Once the phosphoserine group is exposed to the extracellular environment, it is recognized by Annexin V, a protein responsible for the signaling of apoptosis and the consequent activation of phagocytosis.

1.2.1 Annexin V

Annexin V is a 36 KDa single stranded protein that binds PS in a Ca^{2+} dependent manner[13,14] and with a notably high affinity constant with a Kd < 10^{-10} M [15]. From the X-Ray structure of the protein, the binding mode of Annexin V to the glycerophosphoserine group has been determined, where two Ca^{2+} ions are coordinated to the PS headgroup[16]. Annexin V presents specific features for the recognition of PS based on the charge separation and geometrical orientation. Thus, the presence of the other phospholipids does not influence the action of Annexin V in the PS-binding functions (figure 1.3).

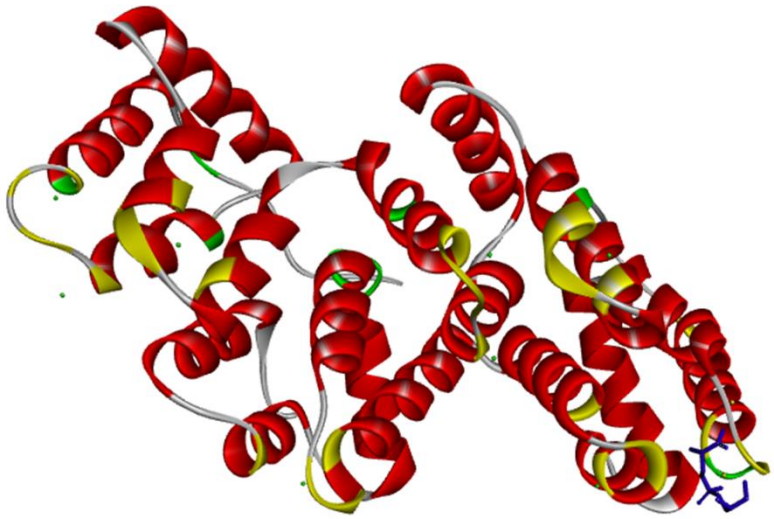


Figure 1.3: Crystal structure of Annexin V (PDB-ID: 1A8A) with its active sites (yellow) and phospholipid head group (blue).

Due to the high affinity shown by annexin towards PS, imaging of apoptosis can be obtained via strategies based on the dye-labeled annexin V as a tool for the assessment of chemo and radiotherapies. Although the annexin V modification for PS detection seems promising, there are drawbacks related to the bioavailability and biostability of the protein. Apart from its large size and high cost, one hour is required for the completion of the annexin binding to PS, which hinders the kinetic assays[17]. In addition, the dependence of annexin V on Ca^{2+} to interact with the headgroup of PS, can lead to false positive results. This can be due to the possibility that scramblases transport PS towards the cellular surface when the concentration of Ca^{2+} increases over the standard level. Moreover, annexin V can also associate with the by-products of lipid peroxidation at the membrane surface[18]. To summarize, an alternative methodology for the detection of PS in vivo is required which should be based on a cheap, low-molecular weight, robust mimic, able to bind rapidly to PS without penetrating the membrane and independent of the presence of Ca^{2+}.

Over the last decade, several research groups have focused on miniature mimics of annexin V as low molecular weight molecules for the detection of apoptosis. Thanks to the study of the X-ray structure of annexin V, the features of the interaction between the headgroup and the protein could be determined. It was observed that the phosphoserine group associates with one of the four binding domains of the protein via two Ca^{2+} [16]. This binding mode lead to simplified models based on the appropriate coordination of two or more metal dications in an optimized scaffold. The miniature receptor should provide the optimal geometry, spatial orientation and charge to interact with the head group of anionic phospholipids preferentially over zwitterionic phospholipids.

1.3 Synthetic receptors for phosphatidylserine

One of the first models of a synthetic receptor for the recognition of phosphates in aqueous media was reported by Hamachi in 2002[19], in which they designed a binding motif for phosphorylated peptide surfaces (figure 1.4). The model was based on the active site of enzymes in which substrates contain a phosphate moiety. Therefore, the artificial receptor possesses a metal complex formed by the coordination of Zn^{2+} in a dipicolylamine pocket for the mimicry of the enzymatic interactions. The design includes an anthracene as a fluorescent moiety for the quantification of the interaction with the phosphate anion. Results showed that the receptor was highly selective towards phosphate anions and presented a remarkable binding affinity, with an association constant of 10^7 M^{-1}. Thus, following the successful results by the group of Hamachi and the increasing interest on the detection of phosphatidylserine, the group of Smith studied thereafter the interaction of the previously reported receptor with the phosphoserine group of PS. This was done to assess the presence of external PS in vesicles and cells by fluorescence microscopy and flow cytometry[20].

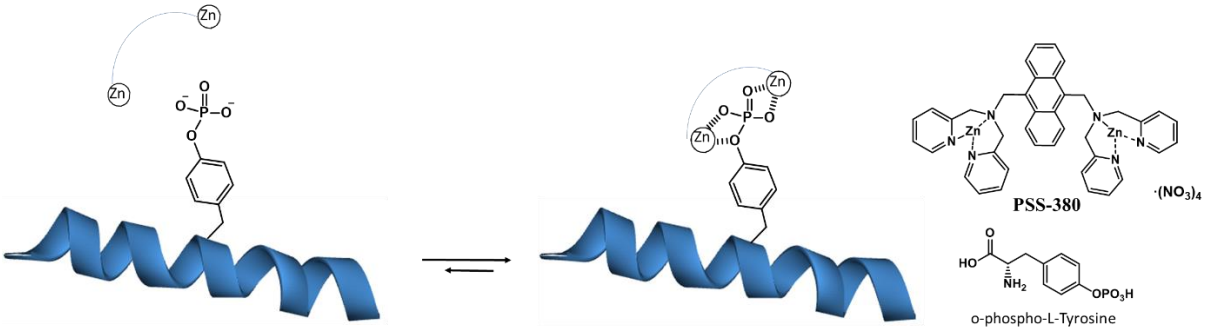


Figure 1.4: Schematic Representation of the interaction of dipicolylamine zinc(II) complex with a phosphorylated peptide.

Although the receptor could be considered an example of PS recognition in cells using synthetic receptors, the anthracene fluorophore suffered from rapid photobleaching, not ideal for time-dependent experiments. In addition, the wavelength at which most common flow cytometers work is outside the UV range, and the wavelength required for the excitation of the anthracene moiety is 380 nm. Therefore, the use of fluorophores that absorb at longer wavelengths was necessary.

Consequently, a second generation strategy was adopted by the same group[21]. The main elements of the receptors were the PS binding unit, a linker and a fluorophore. In order to study the parameters that may influence the interaction of the binding units with PS, linkers with different hydrophilicities were incorporated between the fluorophore and the Zn^{2+} complexes (figure 1.5). The results showed that the tris(ethyleneoxy) linker enhanced the emission of the sensor when binding to PS-enriched membranes. In this case, the Zn^{2+}-DPA affinity units are appended to a phenyl ring, which is linked to the 7-nitrobenz-2-oxa-1,3-diaza-4-yl (NBD) fluorophore. This fluorophore is excited at 470 nm and emits at a broad range around 530 nm. Consequently, its use is limited for the study of mechanistic experiments. Thus, different fluorophores were tested in the design, including fluorescein[22].

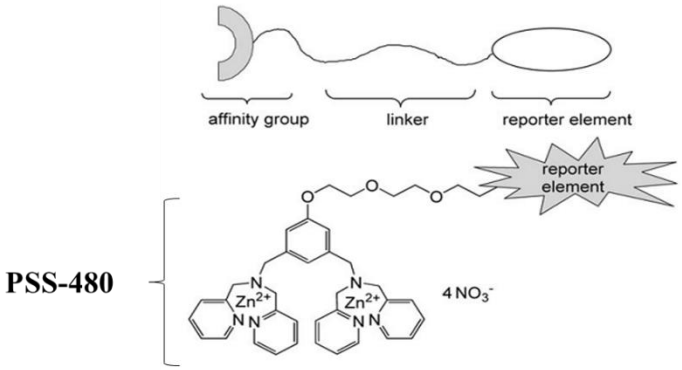


Figure 1.5: Modular design of PS sensor[1].

In order to develop a PS sensor with the same specificity towards PS as displayed by PSS-380 and PSS-480 but with the possibility to detect PS externalization and cell apoptosis, a new series of artificial sensors were designed with enhanced fluorescence properties. The new models were based on the conjugation of the PS affinity group to CdSe/CdS quantum dot system (PSSgreen QD)[22] (figure 1.6). The main advantages of the incorporation of quantum dots into PS detection in cells is the resistance towards photobleaching, enhanced binding due to the multivalency effect and the feasibility of fluorescence microscopy due to the high quantum yield of QDs. On the other hand, extensive washing steps must be done after staining the cells and the high brightness of the QDs could lead to false positive results.

[1] Reproduced with permission from "Hanshaw, R. G.; Lakshmi, C.; Lambert, T. N.; Johnson, J. R.; Smith, B. D. Fluorescent detection of apoptotic cells by using zinc coordination complexes with a selective affinity for membrane surfaces enriched with phosphatidylserine. Chembiochem 2005, 6, 2214–2220".

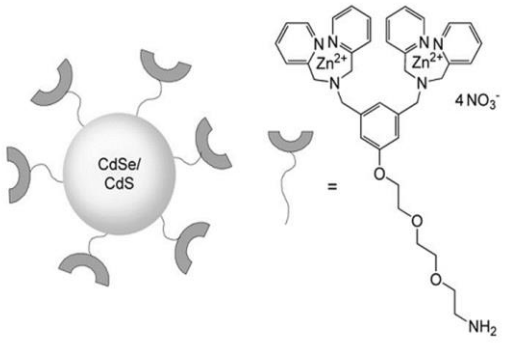


Figure 1.6: CdSe/CdS quantumdot system (PSS-green QD) coated with the PS affinity group[1].

Another design was obtained by exploiting the fact that selective binding to an anionic membrane surface can be achieved by two Zn^{2+}-DPA units with an optimal distance[23]. Thus, a flexible architecture may be beneficial for an optimized PS sensor with two widely spaced binding units (figure 1.7). The obtention of the dipicolylamine units within the peptide structure was achieved by the modification of two ornithine side chains. In contrast to the conclusions gained from previous structural studies[22,24], the results proved that two closely positioned dipicolylamine binding units are not required to achieve PS detection. The affinity constant of the receptor with PS present in artificial vesicles was $5 \cdot 10^4$ M^{-1}.

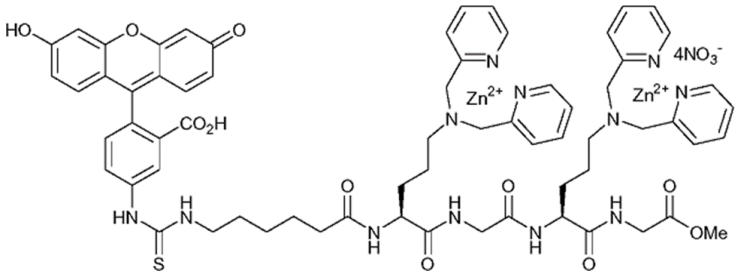
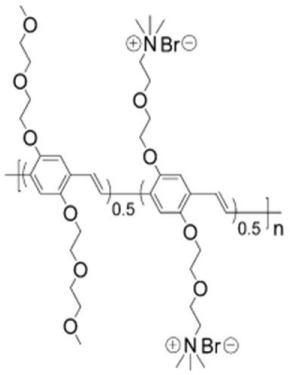


Figure 1.7: Dye-labelled Zn^{2+}–DPA peptide.

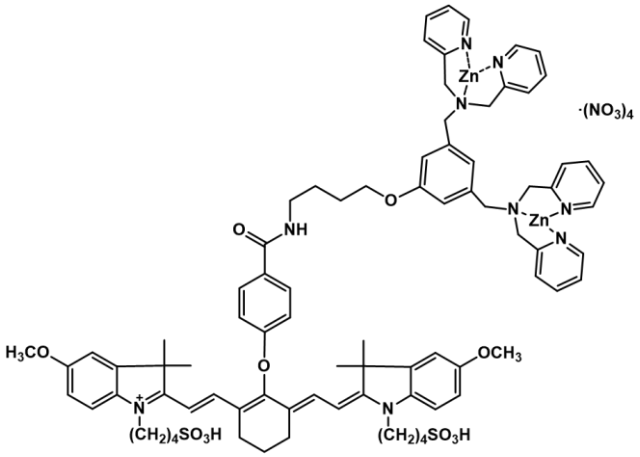
The group of Wang developed in 2011 a new type of novel, inexpensive and label-free PS sensor. It was based on a new water-soluble, positively charged and fluorescent poly(p-phenylene vinylene) derivative (PPV-1)[25] (figure 1.8). The characteristics of the design were the water solubility imposed by the oligo(ethylene glycol) side chains and the self-brightness obtained from the PPV-1 whose absorption range allows the use of fluorescence techniques for the quantification of PS detection.



PPV-1

Figure 1.8. Structure of PPV-1[1].

One of the more recent receptors developed by the group of Bradley D. Smith in 2010[26] contains a near-IR fluorescent probe for the optical imaging of mammary and prostate tumors in living animals (figure 1.9). The near-IR carbocyanine fluorophore absorbs and emits at 794 nm and 810 nm respectively and therefore is applicable for imaging through skin and tissue. Moreover, further studies with this sensor indicated the possibility of preclinical optical detection of tumor cell death due to therapy[27].



PSS-794

Figure 1.9. Chemical structure of PSS-794.

Apart from annexin V, lactadherin is also involved in the recognition of PS in the outer monolayer of the cellular membrane without the need for Zn^{2+} as cofactor (figure 1.10). Thus, mimics of the protein have been developed for the cofactor-free noninvasive detection of PS in cancer cells allowing the straightforward assessment of the efficacy of anticancer drugs[10]. The design developed by the Gao group

[1] Reproduced with permission from "Zhu, C.; Yang, Q.; Liu, L.; Wang, S. A potent fluorescent probe for the detection of cell apoptosis. Chem. Commun. (Camb). 2011, 47, 5524–5526".

focused on the simplification of lactadherin protein. Due to the circular arrangement of the key residues that bind PS in the binding pocket of the protein, Gao suggested a mimic where a cyclic peptide served as a scaffold for the appendage of the recognition residues. In addition, this design did not require the employment of Zn(II) as a cofactor and bind PS with an affinity constant of $1.25 \cdot 10^7$ M^{-1} (figure 1.11). For the design of cLac-2, they linked Arg148 to the rest of the molecule by using two spacers: a dipeptide (-Gly-Gly-, Linker I) bridging Gln85 and Arg148, and a tripeptide (-Gly-Gly-Gly-, Linker II) linking Arg148 and Trp26. Molecular modeling of cLac-2 shows that the two linkers should be sufficiently long and flexible to allow the native-like spatial arrangement of these residues for PS recognition. They also observed that hydrophobicity was needed to partition into the membranes, and therefore they substituted phenylalanine for biphenylalanine.

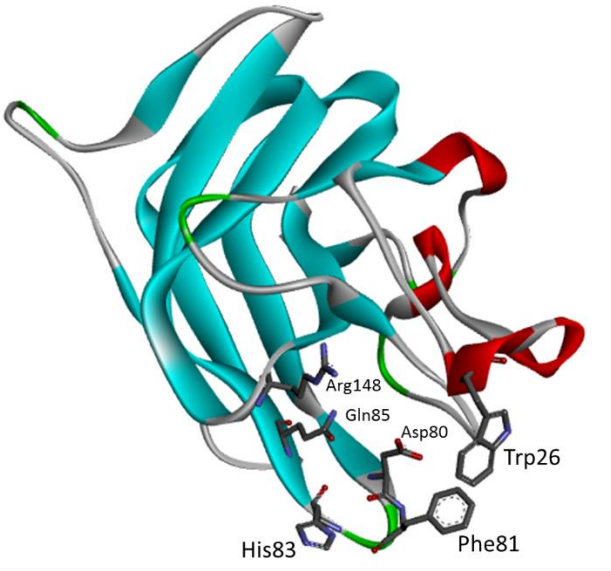


Figure 1.10. Crystal structure of Lact-C2 (PDB: 3BN6) naming the residues that are critical for PS recognition. Hydrophobic residues Phe81 and Trp26 insert into the membrane and residues Asp80, His83, Gln85 and Arg148 that maintain specific polar interactions with the PS headgroup.

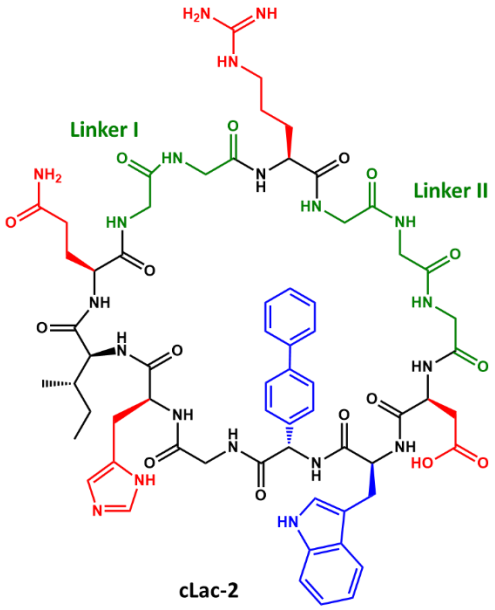


Figure 1.11. Structure of PS-targeting cyclic peptide cLac-2 that incorporates biphenylalanine as a membrane anchor.

Very recently, the Smith group developed a novel PS sensor based on the formation of liposomes as receptor carriers[28] (figure 1.12). The new design combines several advantages of previous PS sensors like multivalency and selectivity. Furthermore, the pharmacokinetic properties of liposomes allow its application on the imaging and theranostic treatment of mammalian cancer cell death, as the liposomes can function as drug delivery carriers[29,30]. In conclusion, the new PS sensor combines cell death targeting and cytotoxic delivery within cancerous tissue.

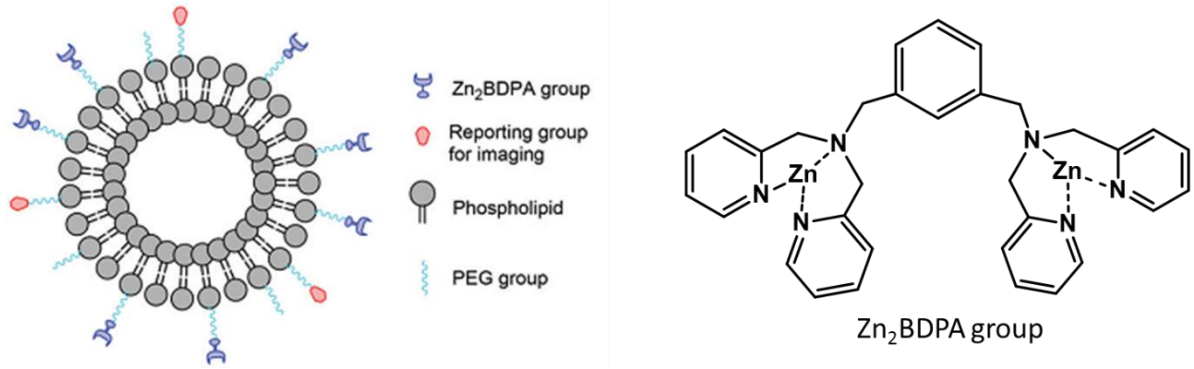


Figure 1.12. Schematic picture of a Zn_2BDPA coated liposome with PEG and Zn_2BDPA groups on the inner leaflet omitted for clarity[1].

[1] Reproduced with permission from "Turkyilmaz, S.; Rice, D. R.; Palumbo, R.; Smith, B. D. Selective recognition of anionic cell membranes using targeted liposomes coated with zinc(ii)-bis(dipicolylamine) affinity units. Org. Biomol. Chem. 2014, 12, 5645–5655".

In our group, we became interested in the development of a new type of receptors for PS in our initial steps towards phosphate recognition. In view of the large expertise of the OBCR group on the decoration of steroids for multiple applications, we have based our research on steroid conjugates as PS receptors. Moreover, the precedents of the groups of Anthony P. Davis and Bradley D. Smith on the modification of bile acids for their use as scramblases for phospholipid transport across the cell membrane uncovered the necessary characteristics for the design of steroid-based receptors to achieve binding of phosphate anions. In addition, we envisaged that cofactors were not mandatory for phosphate binding. The idea of the project, design and synthesis of the receptors is explained in the following section.

1.4 Synthetic receptors based on steroid conjugates

1.4.1. Phosphatidylserine as a target for the mimicking of scramblases

Receptors described in the previous section cover a full range of PS targeting models. As mentioned before, the importance of phosphate anions and phosphate-containing molecules in biological functions has led to the development of synthetic receptors capable of mimicking or modulating the role of natural phosphate recognition elements in an organism. Many processes are regulated by the presence of phosphorylated molecules in the cellular membrane. The asymmetric distribution of phospholipids in the bilayer membrane is a characteristic and sometimes even a requisite for normal cell functions[31]. The natural appearance of phosphatidylserine in the inner part of the membrane is crucial for exocytosis, intracellular fusion processes and signal transduction pathways[32,33]. On the other hand, phospholipid randomization and subsequent externalization of PS to the outer membrane surface is an indication of cell death and clearance by phagocytosis during apoptosis (section 1.1 and 1.2), as well as an activation signal for hemostasis and thrombosis[34]. It has been shown that PS is also targeted by the Alzheimer's amyloid-β-peptide[35]. Therefore, several groups have focused on enzymes that have carried out the translocation of PS across the membrane. These can serve as models to modulate the transport of phospholipids at the outer and inner surfaces with potential applications to cancer therapies and genetic disorders[36]. Research on synthetic scramblases for the recognition and transport of phospholipids has mainly been developed by the group of Anthony P. Davis in collaboration with the group of Bradley D. Smith by exploiting their broad expertise on steroid-based receptors.

In 1990, the group of Prof. Davis began exploring the area of anion recognition. Their starting point was established by the potential of Valinomycin (figure 1.13 A), a natural transmembrane cation carrier, to extract potassium ions from water and transport them across the membrane interior. Unfortunately, no anionic counterpart could be found in nature. Thus, Davis proposed a new design of receptors which could act as potential anion carriers and named them "anti-Valinomycins". It seemed a logical conclusion that cationic receptors should bind better to anions than electroneutral ones. Unfortunately, lipophilic cations were toxic and they required a counter-ion to compensate for their solubility issues, which in turn could generate complications during testing of anion binding capability. The first examples of electroneutral receptors were based on the neutral H-bond donors in the group of Reinhoudt[37,38] (figure 1.13 B). Although

it set up a starting point, the design lacked rigidity and pre-organization. Moreover, intramolecular H-bond interactions formed decreased the interaction with the anion (figure 1.13 C).

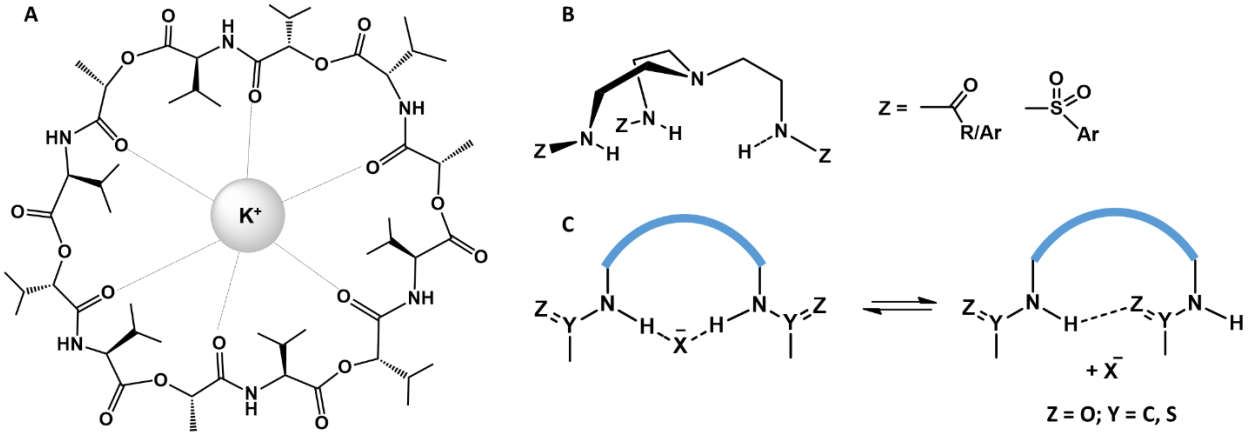


Figure 1.13. A) Valinomycin structure and binding mode to its substrate K^+. B) Anion receptors from the group of Reinhoudt. C) Representation of the intramolecular interactions in electroneutral receptors that can damage the binding site.

1.4.2. Bile acids as templates

Davis then proposed the employment of steroids as rigid frameworks and pre-organized functional group-containing scaffolds for their conversion to H-bond donor arrays. Initial results showed the need for amines instead of alcohols as functional groups for the construction of the binding pocket wherein the phosphates were accommodated. Thus, the synthesis of protected aminosteroids was successfully investigated and reported[39–41]. Furthermore, it was discovered that for a given H-bond donor geometry, steroid-based receptors with more acidic H-bond donors would bind with higher affinity and selectivity. Therefore, electron-withdrawing groups were used to increase the NH acidity, like ureas and thioureas (figure 1.14). At this point, the group of Davis started a collaboration with the group of Smith in order to test the ability of these cholapods as anion transporters interacting with the head-group of phospholipids, such as phosphoserine from PS, through membrane interiors, acting as synthetic flippases[36,42] (figure 1.15). In addition, they have shown that amides and ureas showed preference for phosphates binding over sulphates.

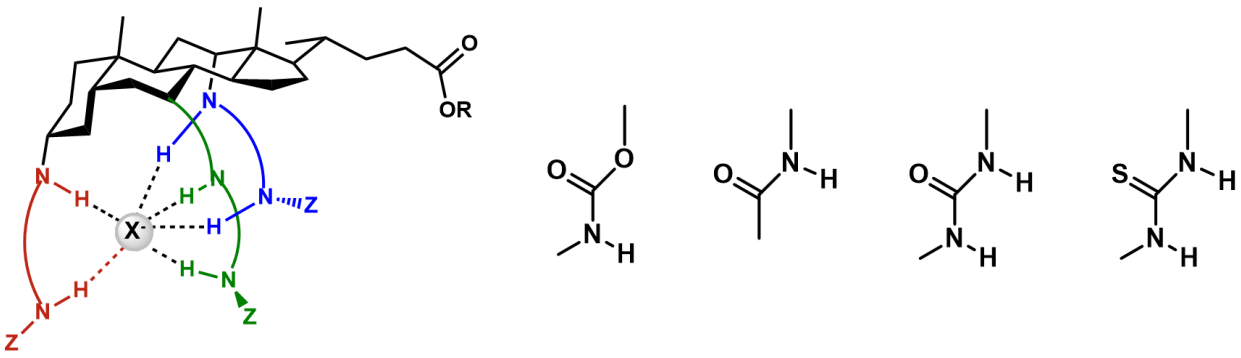


Figure 1.14. Model of neutral anion receptors based on steroid constructs. Z groups adjust NH acidity. On the right, the different functionalities with one or more H-bond donor groups are shown.

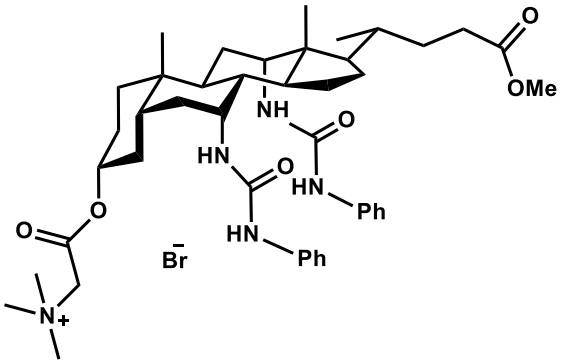


Figure 1.15. Neutral bis(phenylurea)-based receptor developed by Davis.

1.4.3. Bile acids as templates for early detection of apoptosis

Following the success of other groups in the design and synthesis of steroid-based receptors for the recognition of PS, the OBCR group was also encouraged to investigate new types of receptors for the early detection of apoptosis. We here report our first attempt in this direction. The phosphate anion possesses a tetrahedral geometry and in contrast to halogen anions, its extended volume suggested the requirement of a bigger receptor's cavity. We have proposed three different potential receptors of which the cavity can be controlled by cyclization. The acyclic form should show more affinity and selectivity towards phosphate anions due to a higher flexibility for adapting the geometry as is required for the binding of a tetrahedral $H_2PO_4^-$ anion. On the other hand, as the recognition area becomes smaller, the affinity for bulky anions in comparison to halogen anions decreases[43].

In addition, recent studies on the potential of the 1,4-disubstituted 1,2,3-triazole ring to recognize, bind, and sense oxoanions[44] together with the high regioselectivity and efficiency of CuAAC (Copper(I)-Catalyzed Azide Alkyne Cycloaddition) inspired us to incorporate triazoles in the dipodal recognition units amended together by the steroid scaffold. The capacity of triazoles to act as H-bond donors and to participate in dipole-dipole interactions has been already described[45] for anion recognition with a broad variety of receptors[46] and settled for the C–H hydrogen bond donor as a new noncovalent binding motif. It also allowed the modification of peptide structures as amide bond mimics[47], establishing triazoles as new, versatile and successful anion-binding motifs that can be added to the design of receptors, widening the existing list of hydrogen bond donors and Lewis acids already known. Consequently, the group of Pandey has explored 1,2,3-triazolium systems in a bile acid template as an anion-sensing moiety for phosphate anions with the extra benefit of a protonated triazole, enhancing the hydrogen bond donor capacity (figure 1.16). Additionally, he observed that the bridging methylene protons participated in hydrogen bonding with the anions along with the C-5 triazolium protons. Unfortunately, as previously mentioned, the employment of cationic receptors can cause complications.

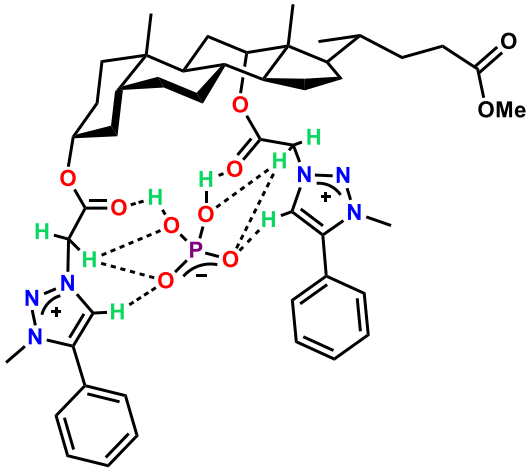
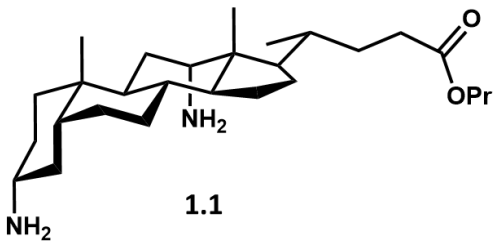


Figure 1.16. Binding mode of Pandey's receptor by H-bond donor interactions with the phosphate anion.

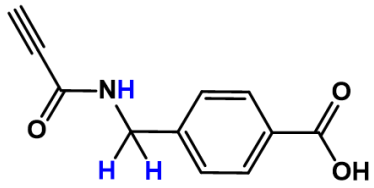
Contrary to the design of Pandey, where he described the appendage of the recognition units by ester bonds, and as illustrated by Davis, all the more promising structures encountered at least one NH group attached to the steroid. We have recently published an optimized strategy directed towards the synthesis of an orthogonally-protected diamine deoxycholic acid, in which one of the intermediates was the 3α,12α-diamino-deoxycholic acid[48] (figure 1.7; compound 1.1).



1.1

Figure 1.17. Intermediate compound in the synthesis of orthogonally protected diamine deoxycholic acid derivatives.

Podands to be incorporated onto the scaffold had to ensure the acidity of the NH group and the presence of an alkyne for CuAAC. Although thioureas have been known as the motif that best increase NH acidity, we proposed amides as the functional group for the linkage of the scaffold with the anion binding motifs as we could overcome the possible decrease in affinity of the amines with the presence of the triazoles. Thus, a carboxylic acid was required, together with the presence of an alkyne. Due to the excellent results obtained in the group of Strömberg for the conjugation of peptides to oligonucleotides by the CuAAC reaction with their taylor-made linker: (N-Propynoylamino)-p-toluic acid (PATA) (1.2) affording an activated triple bond, we proposed its incorporation onto the scaffold as podand to afford H-bond donor units (figure 1.18)[49]. Furthermore, it has been shown that the methylene protons participate in hydrogen bonding.



1.2

Figure 1.18. Structure of the PATA linker **1.2**.

The presence of two alkynes allowed multiple derivatizations towards the final synthetic receptors. The azides were straightforward to obtain by synthetic methods or commercial sources. This ensured the possibility of cyclisation with a bis-azide moiety and the appendage of different units for the synthesis of the acyclic receptors. As an approach towards our goal, we decided to use benzyl azide and hexadecylazide for the acyclic forms, while 1.11-diazido-3,6,9, trioxaundecane was chosen as linker for the cyclization of the podands at both C3 and C12. We here show our first designs for the binding of phosphatidylserine. In addition to having the acyclic form, we have incorporated for the first time an aliphatic chain to create extra hydrophobic interactions with the lipophilic part of PS. As discussed before, the cyclic form should show more affinity for halogen anions. Moreover, publications about artificial receptors for anion recognition based on cholapod scaffolds showed designs with a maximum of 6H-bond donors, providing the strongest binding constants as compared to receptors with less H bond donor capacity. The presence of 4 amide

groups, 2 triazoles and 2 methylene groups should ensure the cooperation of at least 6H-bond donor units in the final recognition cage within the receptors.

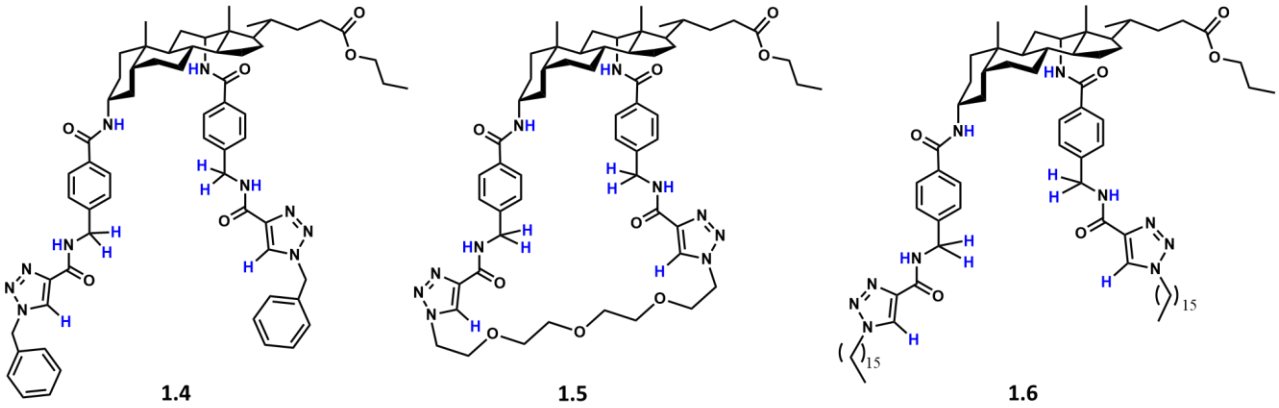


Figure 1.19. Structures of final receptors **1.4, 1.5** and **1.6**.

In order to test the influence of the scaffold on the PS binding, we synthesized an acetylated receptor as a blank. The amine functionalities of compound 1.1 were capped using acetic anhydride and DIPEA to afford **1.7**.

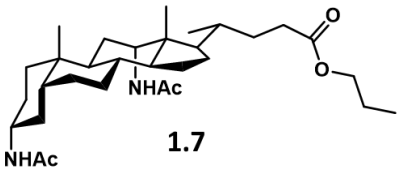


Figure 1.20. Structure of receptor **1.7** used as a blank.

1.5 Synthesis

The diamine scaffold was synthesized according to our published method[48]. The carboxylic acid of PATA is less active for coupling due to the electron-withdrawing effect of the benzyl ring. Therefore the employment of a more reactive coupling reagent such as HATU was required for the amide coupling to the amines. In order to drive the reaction to completion, an excess of PATA was needed. This was probably due to the hindrance present at the C12 position, both because of the structure of the steroid nucleus and the presence of the linker at C3. The C3 conjugation proceeded faster, as the amine at C3 is more exposed to the reaction solution. After purification, the compound was subjected to click reaction with azides, using catalytic amounts of $CuSO_4 \cdot 5H_2O$ and sodium ascorbate under microwave conditions (figure 1.21). Extraction with a saturated solution of EDTA facilitated the removal of Cu particles. Final purification by RP-HPLC afforded the 3 desired receptors.

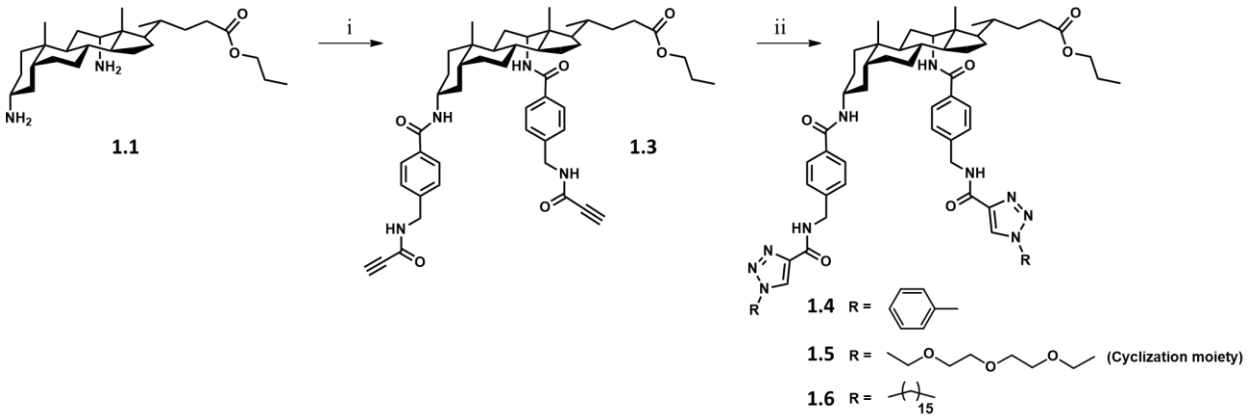


Figure 1.21. Synthesis route for the obtention of final receptors 1.4, 1.5 and 1.6. i) HATU, DIPEA, DMF, 16h, rt. ii) $CuSO_4 \cdot 5H_2O$, NaAsc, 2,6-Lutidine, DMF/H_2O, 1h, 60°C, MW.

1.6 Isothermal Titration Calorimetry (ITC) as a method for the determination of the binding affinity

As a method to study the interaction of receptors **1.4**, **1.5, 1.6** and **1.7** with phosphatidylserine, we chose Isothermal Titration Calorimetry (ITC). The basic principle of ITC can be explained as follows: the determination of the thermodynamic parameters characterizing the interaction in solution (ΔH°, G°, ΔS°, Ka, Kd and n) is achievable by direct measurement of the heat generated or absorbed when the receptor and PS interact. For this purpose, the instrumental setup consist of a reference cell and a sample cell, in which a suitable solvent and a solution of the receptor are placed respectively. A syringe containing a solution of the ligand is inserted into the sample cell and a precise volume of this solution is delivered in the cell at defined time intervals. The interaction is monitored by measuring the heat required (endothermic event) or to be dissipated (exothermic event) to keep the T of the titration cell equal to that of the reference cell after each injection. The heat released decreases after every addition of the ligand due to the diminution of the binding sites available, and it ceases when saturation is reached. The contribution linked to the ligand dilution is determined by injecting the ligand in the solvent used to prepare the receptor solution and subtracted from the heat exchange measured in the presence of the receptor.

In collaboration with Prof. Gilles Bruylants from ULB, we performed ITC measurements of the receptors **1.4, 1.5, 1.6** and **1.7** with phosphatidylserine. Optimization of the conditions (solvents and concentrations) were performed during the course of my PhD. The subsequent work was carried out by Lidia Dumitrescu and Vincent Ornelis. The choice of the solvent was based on the solubility of both receptor and ligand, as they must be both soluble in the same solvent. In addition, the dilution heat of the solvent with the ligand has to be small compared to the interaction heat The concentrations of the receptor and ligand solutions were optimized and found to be 240 and 2400 µM respectively for the receptor (in the cell) and ligand solutions (in the syringe) in $CHCl_3$. Results showed that receptors **1.7, 1.4** and **1.5** do not interact with phosphatidylserine. It was expected that receptor **1.7** would not interact with PS as its structure only consists

on the steroid nucleus, and it was used as a blank to assess the influence of the scaffold on the PS binding. On the other hand, receptor **1.6** showed affinity in the range of 10^4 M^{-1} (figure 1.22), illustrating that aliphatic chains might have an influence on the recognition of PS by ensuring the position of PS head group into the hydrogen bond cavity and by enhancing the interaction via hydrophobic contacts with the aliphatic chains.

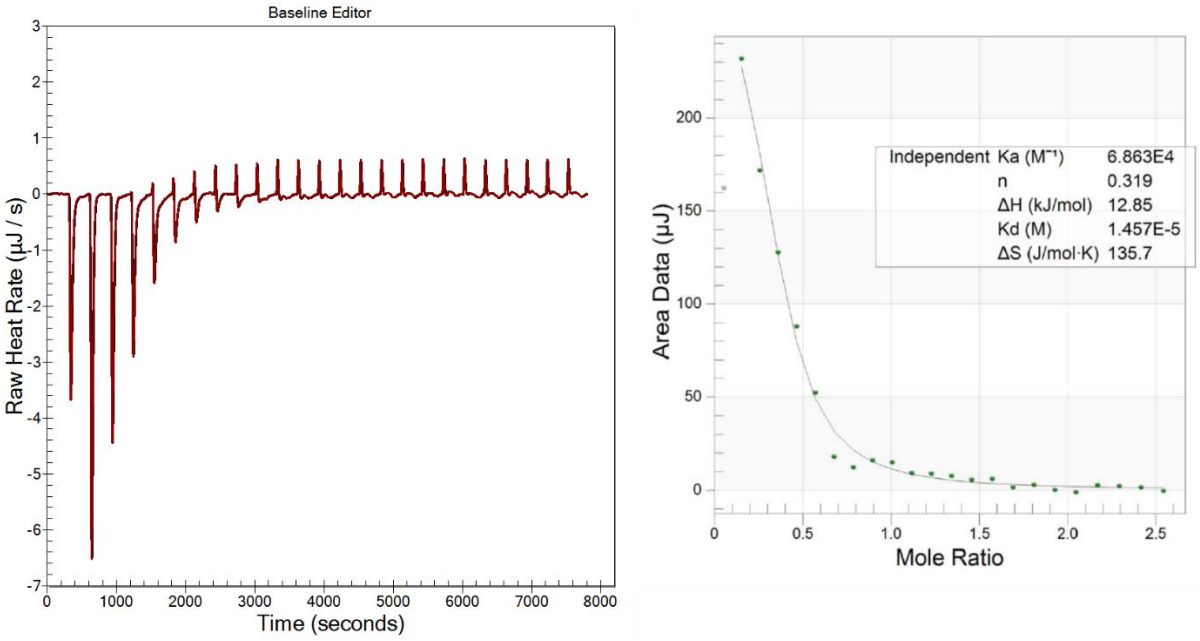


Figure 1.22. Thermogram (left) and integrated data (right) linked to the injection of 10 µl of a 2400 µM PS solution in 1 ml of a 240 µM solution of **1.6**. The profile of the binding isotherm (right) is fitted to the experimental data in order to obtain the thermodynamic parameters.

1.7 Conclusions

In this chapter we present our first steps in the design of bile acid-based receptors for the detection of apoptosis. In addition, we investigated the influence of the size of the binding cavity and the possibility of amide bonds as hydrogen bond donors towards PS. Prior steroid-based receptor[36] in figure 1.14, was tested for the binding of an analogue PS with shorter acyl chains to diminish the aggregation of lipid tails. UV titration technique was used to determine the association constant. The experiments were carried out in 99:1 $CHCl_3$:CH_3OH to simulate the lipophilic interior of the bilayer membrane. The association constant was $(3.4 \pm 0.9) \times 10^5$ M^{-1}. To the best of our knowledge, the best binding constant reported until now is achieved with the cyclic peptide cLac-2 in figure 1.11, and a value of 1.25×10^7 M^{-1}.

In addition, it is known that lipophilic cations like steroid-based receptor in figure 1.14 are toxic and results in solubility problems. Therefore, we present here our first attempts towards the development of neutral

steroid receptors for PS based on hydrogen bond interactions. Moreover, the obtained affinity constant obtained for receptor **1.6** is comparable with the ones obtained with previous reported receptors.

On the other hand, the observed affinity could be due only to hydrophobic interactions between the aliphatic chains from both the receptor and the ligand. In order to discard this option, ITC measurements using as a ligand an aliphatic chain should be performed. Furthermore, the stoichiometry did not correspond to a 1:1 ratio of receptor:ligand. The value n = 0.3 could be due to low purity of the sample measured or by the aggregation of receptors via their lipid tails. In order to investigate if aggregation occurs and if it is necessary for the interaction of PS, further studies are needed in which the concentration of receptor in the sample cell would be reduced to hamper the aggregation effect. Another alternative to assess the interactions within receptors is the employ of Dynamic Light Scattering (DLS) technique, which will allow access to the determination of the size distribution profile of small particles that can be formed by the aggregation of the receptor.

As a conclusion, we have synthesized three receptors for the detection of PS varying the size of the binding cavity. Receptor **1.6** showed an affinity constant in the order of 10^4 M^{-1}. However, further studies are needed to analyze if the interaction is caused by other factors non-related to the hydrogen-bonding between the receptor and the phosphoserine headgroup of PS.

CHAPTER 2

RECOGNITION OF LARGE MOLECULES CONTAINING PHOSPHATES: DEVELOPMENT OF TRANSCRIPTION FACTOR MODELS

Transcription of DNA is one of the most critical processes in biology as it regulates the expression of genes that are needed within the cell to develop essential functions. By targeting DNA in a specific gene, these functions can be modified or inhibited. However, targeting specific DNA sequences requires considerable sophistication in terms of design and synthesis. Mimicking these transcription factors by synthesizing artificial ones gives us valuable insight into the parameters that affect DNA recognition. In this chapter we present our initial attempts on the synthesis of transcription factor models towards the sequence selective recognition of DNA.

2.1 DNA: Discovery, Structure and Function

The elucidation of the Deoxyribonucleic acid (DNA) structure by James Watson and Francis Crick in 1953 was an epoch making discovery in the history of science. Much of our current knowledge with regards to how genetic information is used and stored in a cell is based on research projects that were developed as a consequence of this discovery. Ever since the ground breaking discovery by Watson and Crick scientists all over the world have worked and are still working tirelessly, pouring in their resources to uncover the many mysteries that surround DNA.

To explain in the simplest possible terms the structure of DNA it can be said that each DNA molecule consists of two antiparallel coiled strands around each other to form a double helix, a structure like a spiral ladder. Each rung of the ladder consists of a pair of chemical groups called bases. There are four types of bases – Adenine (A), Guanine (G), Thymine (T) and Cystosine (C) which combine in specific pairs and stabilize each other via hydrogen bonding (A-T, G-C) so that the sequence on one strand of the double helix is complementary to that on the other. It is this specific sequence of bases which constitutes genetic information.

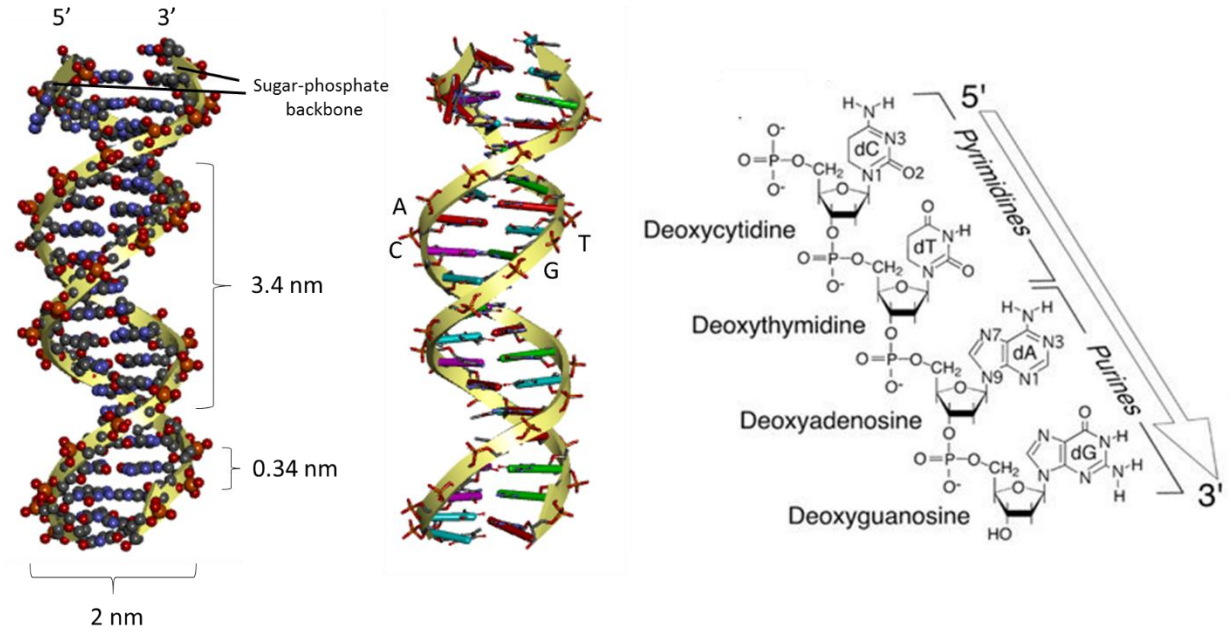


Figure 2.1: Structure of dsDNA: base pairs & phosphate backbone[1] [50].

In more complicated but structurally and chemically correct terms, DNA is an oligomeric molecule formed by the nucleotide units A, G, C, T. These nucleotide units are comprised of a phosphate and a sugar molecule. As a chemist, one can explain the connectivity and define the bonds thus: The β-D-2'-desoxiribose is joined by a β-glycosidic $C_{1'}$-N bond to a nitrogen base and by a phosphodiester bond to the phosphate group at both 3'-hydroxyl and 5'-hydroxyl groups. The phosphates of DNA are quite acidic (the pKa is nearly 0!) and therefore nucleic acids are polyanions at physiological pH = 7.365.

The bases are organized in the form of a helix due to the geometry of the sugar and the hydrophobicity of the bases. Their natural tendency is to minimize the contact surface with water. Moreover, as phosphates are negatively charged, they are stabilized in contact with water. The most common conformation of DNA under physiological conditions is the B form in which both polynucleotide chains are wound in a right-handed conformation around a common axis. The final diameter is approximately 20 Å and the turn of the helix is 36° for each base pair. Thus, a complete 360° turn consists of 10 base pairs with the pitch of the helix measured as 34 Å with 3.4 Å as the length of each nucleotide. The planes of the base pairs are almost perpendicular to the axis of the double helix. B-DNA contains 2 grooves of different widths, the major groove and the minor groove, due to the asymmetrical attachment of the base pairs to the sugar rings of the backbone. The latter one is deep and narrow while the former is wide.

So what is the purpose of DNA? The truth is nobody knows for sure the full range of functions that DNA can perform. Nature chose DNA for a reason – its stability, easy enzyme manipulation, its perfect code

[1] Figure 2.1 (right): Chapter 1: DNA Structure: Alphabet Soup for the Cellular Soul. © 2011 P. Shing Ho, Megan Carter. Originally published in "Biochemistry, Genetics and Molecular Biology" » "DNA Replication-Current Advances" under CC BY-NC-SA 3.0 license. Available from: http://dx.doi.org/10.5772/18536.

which can store massive amounts of information. But the truth is, despite over 60 years of research after its discovery, we have only scratched the surface. This is hardly surprising since it took millions of years for life to evolve. However, what we do know for sure is that in humans, proteins are synthesized based on the expression of information stored in DNA during the transcription and translation processes making it vital that we understand the intricate processes that surround this giant puzzle.

2.2. Transcription: From DNA to RNA

As mentioned before, it has been concluded after years of research that indeed the four different nucleotide subunits are responsible for the encoding of genetic information which are then stored in sequences that make up DNA. Further developments in cell biology and biochemistry have led to the revelation that this genetic information is in turn responsible for cell functions.

Cells comprise of several things – ranging from sugars to lipids to DNA. Proteins are still, however, one of the main components in the cells and determine its structure and function. The synthesis of proteins is dictated by the DNA sequence and its encoding process through replication, transcription and translation. This process ensures transferring the genetic information from the "language" of the nucleotides to the "language" of the amino acids.

The synthesis of proteins is a complex process in which DNA is the keystone. Once a protein is needed in the cell, the nucleotide sequence corresponding to it is transcribed into RNA by an enzyme, RNA polymerase II. Each DNA segment containing the required information for the synthesis of a protein is called gene. Each gene is thus translated into a protein as and when required. These are the fundamental concepts involved in the expression of genetic information and therefore constitute the central dogma of molecular biology.

As part of this thesis, we will focus mainly on the process of transcription. The first step in gene expression is to copy the sequence of a specific gene into RNA. This process is called transcription due to the almost similar nature of the main elements involved, nucleotides. RNA, as DNA, is a polymer constituted by four different nucleotides linked by phosphodiester bonds. Although DNA and RNA have similarities, the differences between their structures result in important consequences in terms of the functions they perform. RNA, unlike DNA, is single-stranded and therefore can adopt a broad range of 3 dimensional structures. This property of folding into different conformations confers upon the RNA the ability to act not only as an information storage, but also as a catalyst. In addition, RNA has can also have structural and regulatory functions in the cell.

During transcription, RNA is synthesized from a segment of double-stranded DNA. The process is initiated by exposing one of the DNA strands by unwinding the double-stranded DNA of the desired sequence. This allows the RNA polymerase II to recognize and bind to the promoter sequence of the gene to be transcribed. The enzyme can only bind to its sequence through the presence of certain proteins, transcription factors to be precise, that bind selectively to the promoter site and form the transcription initiation complex. One of

the two strands serves as a template for the complementary base-pairing synthesis of an antiparallel RNA strand. In layman terms these transcription factors "read" the genetic code which is built into dsDNA.

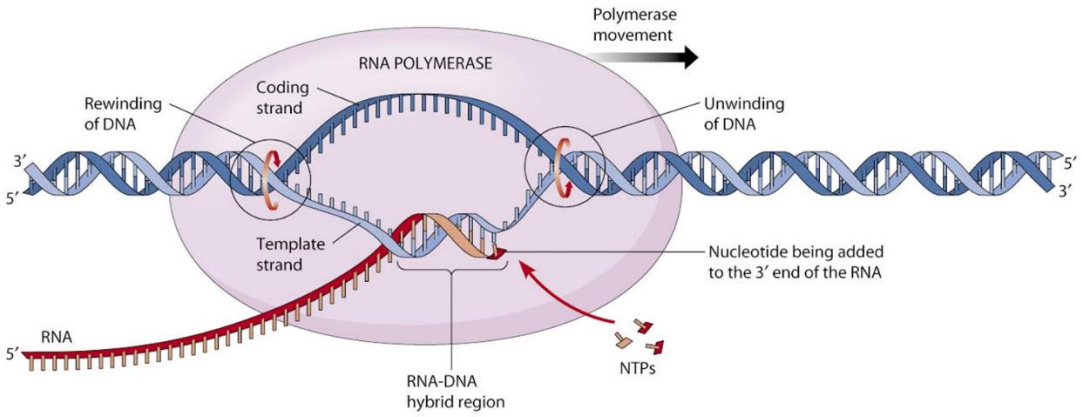


Figure 2.2. Representation of the DNA transcription process[1].

As RNA polymerase II synthesizes RNA in the 5' to 3' direction, the complementary DNA strand used as the template is the one oriented in the 3' to 5' direction. Ribonucleotides are added covalently to the growing RNA chain by phosphodiester bonds while the enzyme moves through the DNA template strand unwinding the helix and exposing the DNA sequence to complementary base-pairing with the incoming ribonucleotides. Elongation of the RNA chain is terminated when the RNA polymerase encounters a specific DNA sequence, where the enzyme loses its affinity for the DNA. The newly synthesized RNA chain is then released and ready for translation or post-transcriptional functions.

2.3 DNA transcription as a key process for gene therapy

Due to several complex mechanisms involved and a lack of clear understanding and knowledge about the specificity of the interactions between DNA and proteins the study of the regulation of gene expression has always been of particular interest. Moreover, this lack of knowledge has significantly hampered the development of gene therapy strategies in the past and hence study of the interactions between proteins and DNA has gained a lot of traction in recent years especially where other therapeutic strategies fail. In some diseases, particular genes are overexpressed causing interferences with biological functions in the cell. Therefore, the prevention of specific genes from being transcribed could ultimately result in a therapeutic platform with biomedical applications[51]. During the past decade, biologists and medicinal chemists have been working together in order to target specific genes for this purpose. In nature, some proteins interact with DNA in a sequence-specific way by means of their recognition domains, which consists of a short sequence of amino acids. Simplification of a protein could ultimately lead to a peptide-based drug in cancer treatment. Moreover, due to the improvements in peptide manufacturing, peptide drugs can be obtained through synthetic methods and techniques which allow the increase of peptide stabilization. In this context,

[1] Reproduced from http://limbiclab.com/.

Transcription Factors (TFs) have been studied in great detail for the modification of gene expression and the study of DNA binding affinity in the cellular environment.

2.4 Interactions between DNA and proteins: The key factor regulating transcription

Transcription factors regulate transcription in eukaryotic cells by determining not only the genes to be transcribed but also the rate of transcription. They bind and recognize regulatory DNA sequences related to the gene (figure 2.3). The regulation is performed via very specific interactions between the TF with the regulatory gene of the target gene by activating or repressing its transcription.

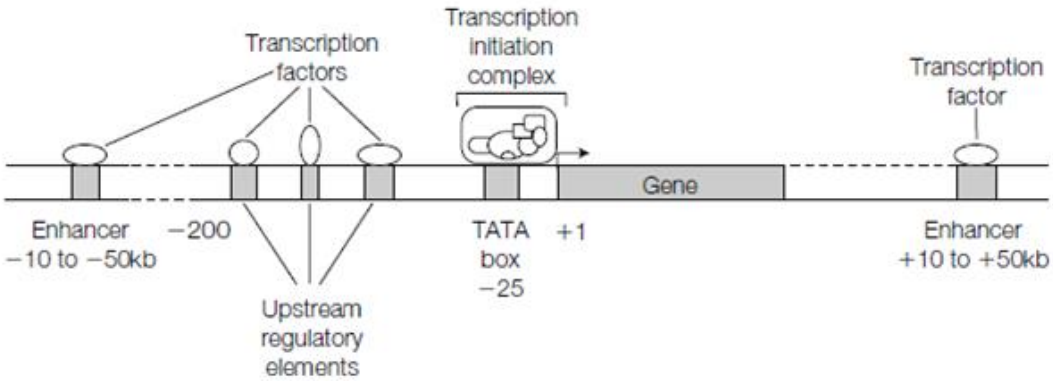


Figure 2.3. Control regions that regulate transcription of a typical eukaryotic protein-coding gene. Although shown as distinct entities in the figure for clarity, in vivo the different regulatory proteins bound to the control elements and distant enhancers interact with each other and with the general transcription factors of the transcription initiation complex to modulate the rate of transcriptional initiation[1].

2.5 Types of interaction between proteins and DNA

The lack of a general recognition code for the interaction between amino acid sequences within a protein and its specific DNA-binding site has promoted the study of the structure of TFs and its interaction with the DNA. Although a general formulation of these interactions has not been achieved, some general rules can be followed, providing a starting point for the molecular basis of specific DNA recognition.

Due to the fact that the secondary structures in case of proteins are extremely well defined, very specific interactions take place between the amino acids of proteins and the complementary base pairs of DNA. In case of TFs, all these interactions occur in the major groove of the DNA, which is wide enough to accommodate the α-helix recognition unit. A detailed explanation of the four major forces between proteins and nucleic acids can be summarized as follows:

1. Electrostatic forces: Salt bridges. These are long range, electrostatic interactions between groups of opposite charges. In this case, these are quite important in view of the presence of the negatively charged DNA due to the phosphodiester backbone and the positively charged

[1] Reproduced with permission from: http://www.expertsmind.com/topic/regulation-of-transcription-by-rna/mechanism-of-regulation-95614.aspx

basic side chains of the amino acids of the protein, such as Lys, Arg and His. These interactions contribute to the free energy of association and do not influence the specificity.

2. Dipolar forces: Hydrogen bonds. These can be defined as short range interactions that contribute substantially to the specificity of the interactions that occur between proteins and DNA, however, they do not provide stability to the complex. They occur between the polar side chains in the proteins and the sugars and bases in the DNA. Higher degrees of specificity can be achieved by bidentate hydrogen bonds between a single side chain and a base or base pair.
3. Entropic forces: The hydrophobic effect. These are short range, non-directional forces and contribute to the free energy of association. These non-polar contacts occur between the apolar amino acid side chains and the DNA base pairs.
4. Water-mediated hydrogen bonds. Water molecules behave as space fillers in the protein-DNA complexes and influence the positions of hydrogen bond donors and acceptors at the base edges.

2.6 Families of transcription factors

The protein structure of TFs can be divided into two broad domains: the DNA binding domain which is the part of the protein responsible for the selective interaction with the DNA control sequence for transcription. The other domain is the DNA activation domain, the part responsible for the interaction with other TFs or with the RNA polymerase II.

There are several DNA binding domains found in TFs in eukaryotic cells: Helix-turn-helix, zinc finger and the basic domain.

2.6.1 The Helix-Turn-Helix motif:

It consists of two α-helices brought together by a β-turn formed by a short peptide sequence. The two helices are held together at a fixed angle through interactions between them. The C-terminus of the protein is inserted in the major groove of the DNA. Sequence-specific recognition is achieved via the interactions between the side-chains of the amino acids with the DNA which vary between proteins (figure 2.4).

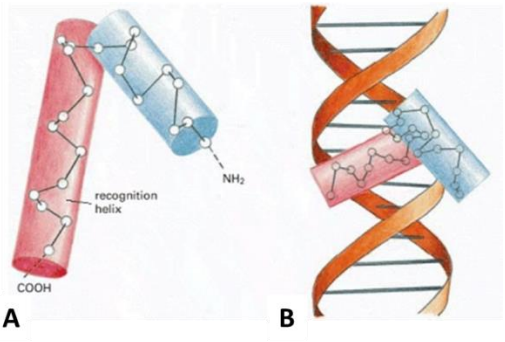


Figure 2.4: The DNA-binding helix-turn-helix motif[2]. The motif is shown in (A). The recognition helix (red) is located at the C-terminal part of the protein. As shown in (B), this helix binds the major groove of DNA, where it is in contact with the edges of the base pairs. The N-terminal α-helix (blue) functions basicaly as a structural component that allows the optimal position of the recognition helix.

The binding mode of the helix-turn-helix proteins is very similar to other sequence-specific DNA binding proteins. HTH proteins recognize double stranded DNA and arrange themselves symmetrically with respect to the DNA sequences having two similar half-sites. This binding feature allows each monomer of the protein to interact with one of the half-sites independently and therefore increase the binding affinity, by increasing DNA contact interface and the helicity conformation of the peptides.

2.6.2 Zinc-finger motifs

The previous helix-turn-helix motif is formed only by amino acids. Zinc-finger motifs present several zinc atoms within their structures. There are two main zinc-finger motifs in this family of TFs: the C_4 finger and the C_2H_2 finger. The latter one presents one α-helix and two β-strands, held together in a compact structure by a loop (figure 2.5). It contains 12 amino acids including two cysteines and two histidines, needed for the coordination of a zinc ion within a tetrahedral geometry. The α-helix region of the protein is responsible for the interactions with the major groove of the DNA. The C_4 zinc finger motif has the same structural arrangement as the C_2H_2 but unlike the C_2H_2, the C_4 has 4 cysteines which coordinate the zinc instead of two cysteines and two histidines.

[2] Adapted from DNA-Binding Motifs in Gene Regulatory Proteins Cover of Molecular Biology of the Cell Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

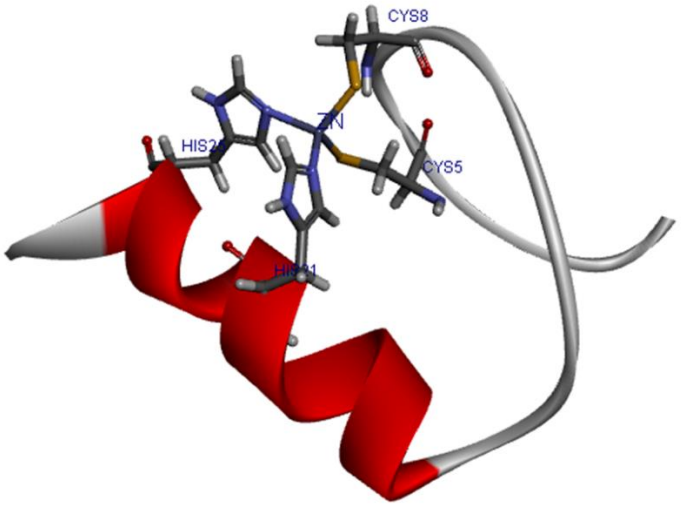


Figure 2.5: One type of zinc finger protein. This protein belongs to the Cys-Cys-His-His family of zinc finger proteins, named after the amino acids that contact the zinc. The three-dimensional structure of this type of zinc finger is constructed from an antiparallel β sheet followed by an α-helix. The four amino acids that bind the zinc hold one end of the α-helix firmly to one end of the β sheet.

In this motif, the strength and specificity of the DNA-protein interactions has been modulated considerably by nature in terms of the number of basic structural units. This can be concluded from the fact that this motif is found as a cluster of zinc fingers and its α-helices can interact with the major groove.

2.6.3 Transcription Factor Domains

2.6.3a Basic domains: Proteins containing basic domains at the N-terminus are rich in basic amino acids (Arg, Lys & His) and possess dimerization domains such as the basic Leucine zipper (bZip) or the basic Helix-Loop-Helix (bHLH) which explains how the names of these TFs have been derived. These proteins bind dsDNA as dimers of two basic domains. Dimerization between monomers occurs due to the interactions between the side chains of hydrophobic amino acids at the dimerization domain, and allows the basic domain to dimerize and grip the major groove of the DNA.

2.6.3b Dimerization domains:

I. Leucine zippers

The dimerization domain of bZIP consists of an α-helix with leucines on the same side of the helix every seven amino acids (i.e. The 1^{st}, 8^{th}, 15^{th}… amino acids will be on the same side). This helps creating a hydrophobic surface and allows the dimerization of the monomers by the interactions within the Leucine zipper motif. Consequently, the leucine zipper enables the basic region monomers to face opposite directions and wrap around the major groove at palindromic DNA sequences (figure 2.6).

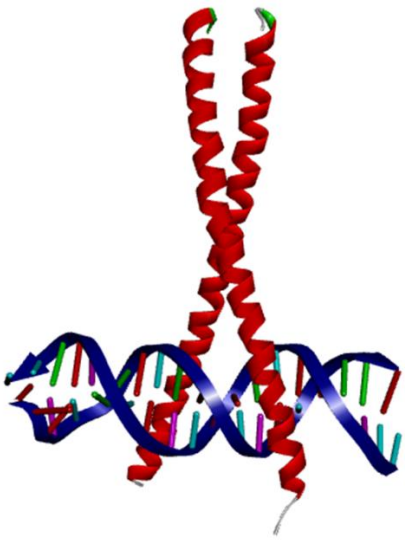


Figure 2.6: A leucine zipper dimer bound to DNA.

II. Helix-loop-helix motif

The main feature of this motif, which makes it different from other leucine zipper motifs, is that the monomers contain a non-helical loop at the interface between the dimerization and the binding domains. The C-terminus of these helices consists mainly of hydrophobic amino acids whose side chains face one side of the helix allowing hydrophobic interactions between the C-termini of the helices thus forming the dimerization region (figure 2.7).

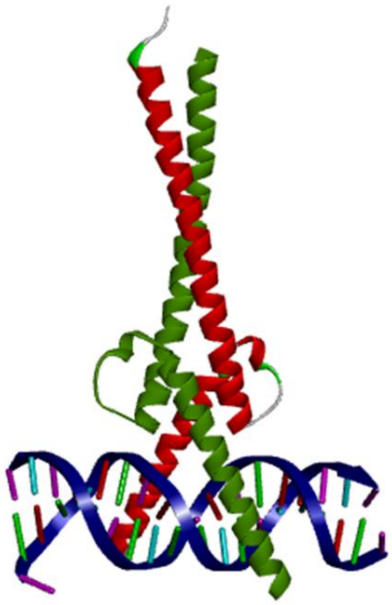


Figure 2.7: A helix-loop-helix dimer bound to DNA.

In case of both dimerization motifs, viz. the leucine zippers and the HLHs, the dimerization can occur between monomers with different amino acid sequences, forming heterodimers instead of homodimers.

2.7 Description of the bZIP Leucine zipper GCN4

The yeast transcriptional activator GCN4 presents one of the simplest protein structures as compared to other TFs present in eukaryotes and it controls the activation of several genes in response to amino acid starvation. GCN4 binds DNA in a sequence-specific manner and its function is the transcription from promoters containing the sequence 5' TGACTC 3'[52]. Its crystal structure was determined by Ellenberger in 1992 by X-ray crystallography with a 2.9 Å resolution[53], available at the Protein Data Bank (PDB-ID: 1YSA) (figure 2.8). In general, this TF is geometrically in the shape of a pair of α-helices that form a parallel coiled coil that diverges as one moves from the C-terminus to the N-terminus ensuring the proper dimer formation and allowing the wrapping of the DNA major groove by its binding domain. The main contacts of this region of the protein with the DNA are established via interactions between the bases and phosphate groups present in DNA and the basic residues present in the protein.

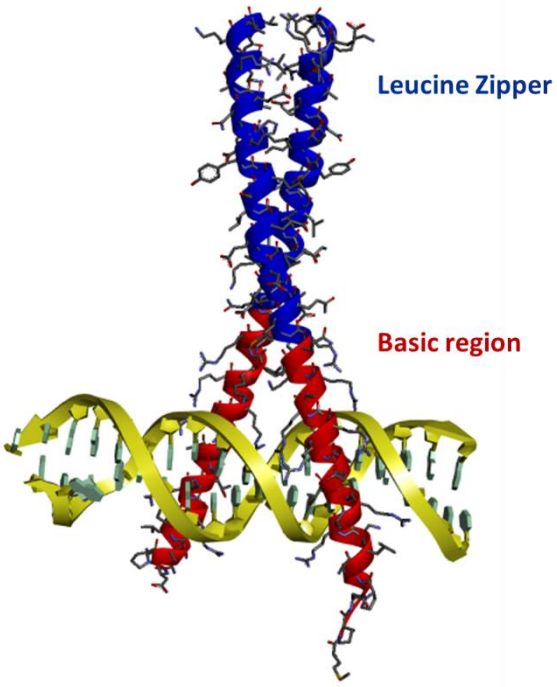


Figure 2.8: Molecular visualization of the GCN4 bZIP Leucine zipper TF bound to the CRE recognition sequence.

The leucine zipper stabilizes the coiled-coil dimer due to the hydrophobic and non-polar interactions within the monomer forming an extensive van der Waals contact surface. Intra and inter-helical salt bridges are present in a low quantity, however, they also contribute to the formation of the dimer.

The basic region of each monomer adopts the optimal conformation to bind to its respective half site of the sequence due to the rigid coiled coil, that imposes a bend of the α-helix towards the axis of the DNA, and therefore grip the major groove and interact with the base pairs located there.

The DNA binding affinity and specificity of this family of TFs is determined by its bZIP DNA binding domain that comprises of 55 to 65 residues, 25 of which are located at the N-terminus and pass through the major groove of each half site of DNA. The main residues of the protein involved in the DNA recognition comprise of the amino acids numbered 226-248 (DPAALKRARNTEAARRSRARKLQ) of the basic region of the GCN4 protein. Numerous contacts of this region with the major groove are based on the positively charged polar residues, which establish hydrogen bonds with the unesterified oxygen atoms of the phosphate backbone and the base pairs. The N-terminal end of the basic region continues as a helix beyond the point of DNA contact, giving the protein an appearance of a pair of α-helical tweezers.

2.8 Artificial dimerization of the bZIP Leucine Zipper basic regions

Due to the simplicity of the structure of bZIP leucine zipper TF described above, miniaturization of this protein into non-natural versions using chemistry-based synthetic routes has been the target of several research groups. The miniature versions should maintain the geometry and the properties of the original protein. The first breakthrough was obtained by Kim's group[54]. They reported a simple yet efficient disulfide bond as a moiety for the substitution of the leucine zipper domain of the GCN4 TF without substantial effect on the DNA affinity and selectivity compared to the natural binding domain.

Other dimerizing units were developed following the basis of Kim's construct (figure 2.9). One of them is the coordination complex described by Schepartz's group[55] in which the two basic regions were connected through a bis(terpyridyl)iron(II) complex. The assembling of the recognition units was performed through the cis and trans conformation of the complex thereby obtaining different geometries which could in turn be used for studying the effect of the orientation of the recognition units on DNA binding.

In addition, dimerization of the monomers by non-covalent constructs was achieved by the Morii group[56]. The model was based on the inclusion complex formation of an adamantane and a cyclodextrin, each of which was conjugated to a single strand of the GCN4 basic region. It was observed that each strand did not bind by itself but the formation of the non-covalent complex resulted in a dimer of the two strands which could then bind DNA. The synthesis of this design extended the knowledge about the importance of the dimer formation. Following these models, the group of Mascareñas studied the modulation of the DNA-binding affinity using a photoswitchable azobenzene linker as dimerization unit between the GCN4 basic regions[57].

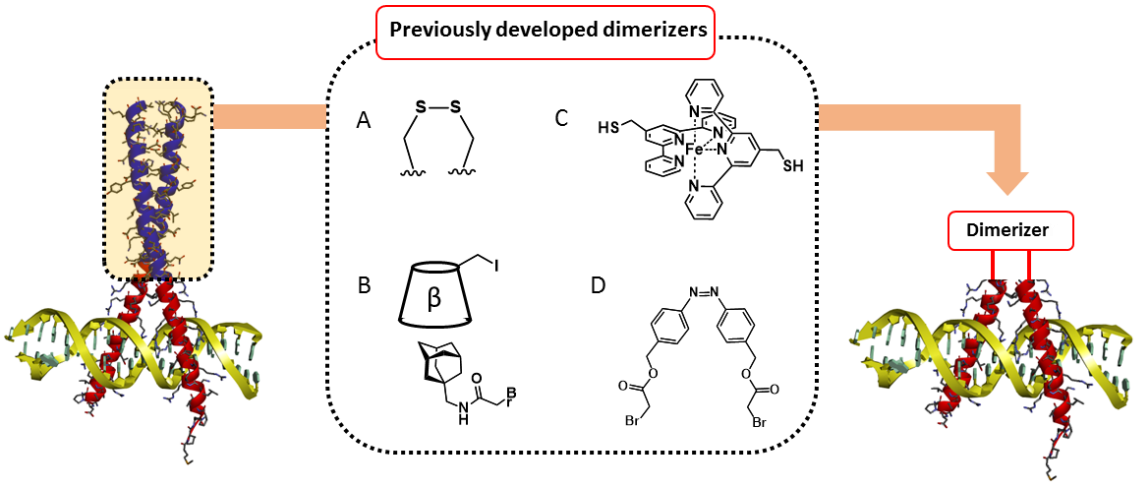


Figure 2.9: Previous scaffolds as dimerizers of GCN4 basic region peptides. A) Disulfide bond, Kim et al[54]. B) Cyclodextrin-adamantane complex by Morii[56]. C) Iron complex developed by Schepartz and co-workers[55]. D) Diazobenzene unit by Mascareñas[57].

Recently, in our own OBCR group, we have reported a dipodal peptidosteroid mimic of the GCN4 TF[58]. The synthesis was based on the linear coupling of amino acids comprising the basic region to a steroid nucleus via SPPS. Although the purity of the final compound was low, it showed DNA recognition properties. A comparison of our previous model to this one is discussed in the following section.

2.8.1 Prelude to the work of the current project: Bile acid as a dimerizer

Over the last decade, steroidal structures have gained significant attention in fields such as supramolecular and bioorganic chemistry, pharmacology and nanotechnology.

Cholesterol is metabolized in the liver and produces bile acids as end products. Thereafter, the steroid nucleus undergoes several hydroxylations one of which includes the loss of an isopropyl group from the side chain, to produce primary bile acids. Further chemical alterations produce secondary bile acids, such as deoxycholic acid.

Bile acids feature several structural differences but those present in higher vertebrates share the same structural skeleton. They consist of two units, a rigid steroid nucleus and a short aliphatic side chain. The steroid nucleus is a tetracyclic hydrocarbon perhydrocyclopentanophenanthrene unit, composed of three 6-membered rings (A, B and C) and a 5-membered ring (D). They contain two angular methyl groups at positions C18 and C19. The A and B rings are in a cis configuration which allows the nucleus to adopt a curved shape. It must be noted that the trans configuration is also possible, but not of interest for the current thesis. Due to the arrangement of the hydroxyl groups in the core structure, bile acids possess an amphiphilic character with a convex hydrophobic upper and a concave hydrophilic lower side (figure 2.10).

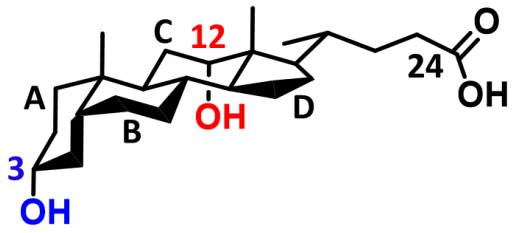


Figure 2.10: Structure of chosen scaffold for current work.

The different chemical properties of the hydroxyl groups on the steroid nucleus also result in bile acids having individual physical and chemical properties that are different from the other members of the bile acid family possessing the same sterane core. This makes them interesting building blocks for the design of novel antibiotics[59,60], anion and cation receptors[61–63], artificial ion channels[64], scaffold for combinatorial chemistry[65] and drug vehicles[66] to name a few among several others.

The specific choice of the steroid scaffold as a dimerizing unit for TF models was inspired by several reasons such as its known ability to enhance the proteolytic stability of the attached peptides[67–69], its amphiphilicity[70], its conformational properties ensuring correct positioning of the two appended chains[71] and its potential to increase cellular uptake and bioavailability[68].

In our specific case, we needed to ensure that the scaffold can perform a function similar to that of the leucine zipper, induces helicity to the basic region next to ensuring the correct distance between the peptides. We chose deoxycholic acid as a starting template for the attachment of the DNA-binding peptides as we envisaged it could perform this function perfectly[72]. Deoxycholic acid contains two hydroxyl groups at the C3 and C12 positions in an α-stereochemistry and can provide a rigid framework for the directional orientation of the peptides. The distance between C3 and C12 is 7-8 Å[73], which ensures sufficient distance for the anchoring of peptides which should then perform an insertion into the DNA units of the major groove. From a synthetic point of view, deoxycholic acid allows the synthesis of peptides via a solid-phase strategy due to the carboxylic acid functional group at C24. This can be connected to a solid support leaving the C3 and C12 positions for peptide synthesis. Synthesis on solid phase allows for an easy purification by simply washing the resin after a coupling reaction.

The difficult step in the whole process was the design of a derivative of deoxycholic acid in order to obtain an orthogonally protected Alloc, Boc-diamino scaffold for further decoration via amide bonds[48,74] (figure 2.11). Esterification of deoxycholic acid at the natural hydroxyl functions was not suited for automated solid-phase procedures, due to the slow rate of the reaction. Moreover, esterification reaction conditions are neither suited nor optimized for automatic synthesizers thus far. In addition, the intrinsic lability of the ester moiety could possibly prevent further elaboration of extended or complex peptide assemblies since repetitive treatment with peptide coupling reagents could result in premature cleavage and/or side reactions. Therefore, emphasis was put on the replacement of the hydroxyl by amino functionalities for the synthesis of peptidosteroid conjugates[75–77].

The first steps towards the synthesis of diamine scaffolds were performed by the group of Davis in the 90s. They performed the conversion of cholic acid to 3α-NHAlloc, 12αNHBoc-cholate in 11 steps with an overall yield of 27 %. The synthesis started with the protection of the carboxylic acid. Prior inversion of the hydroxyl group at C3 allowed the selective oxidation of the hydroxyl group at C12 to a ketone, which was followed by the re-inversion of the alcohol at C3. Oxyme formation and further two step reduction (first with H_2/PtO_2 and then with Zn/AcOH) led to the boc protected amine. Mesylation of the OH at C3 followed by nucleophilic substitution with NaN_3 allowed the conversion to amine by final reduction of the azide with H_2/PtO_2. Protection of the amine with allyl chloroformate allowed the obtention of the desired 3-NHAlloc, 12-NHBoc steroid scaffold.

In our group, a shorter and higher yielding synthetic route towards an orthogonally protected diamine scaffold was developed during the PhD work of Dieter Verzele[48,78], together with the first attempts in the OBCR group on the miniaturization of TFs. This started in 2007 with the simplification of a member of the helix-loop-helix zipper (b-HLH-ZIP) family, the cMyc-Max oncoprotein[79]. For this purpose, he used deoxycholic acid as a template. The synthesis was initiated by the oxidation of alcohols at C3 and C12 to the formation of the corresponding diketone. Oxime formation and further reduction with sodium[76] in n-PrOH afforded the diamine deoxycholate scaffold. Further selective alloc protection of the amine at C3 required tedious purification steps prior to the protection of the amine at C12 with boc. Final hydrolysis of the propyl ester at C24 generated the desired scaffold (figure 2.11)

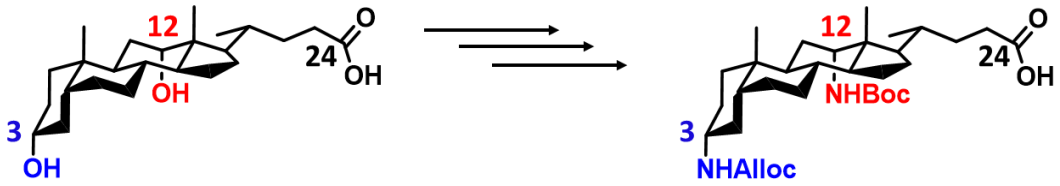


Figure 2.11: Double modification of the C3 and C12 to obtain a diamino derivative of cholic acid. Modification.

The conjugation of the basic region peptides to a steroid nucleus was achieved, but the final compound showed no DNA binding properties. Molecular visualization of the protein confirmed that the direct attachment of the α-helices led to an abrupt truncation of the construct. As a result of this, no insertion of both peptides simultaneously into the major groove could take place due to the rigidity of the conjugate which in turn hampered α-helix formation. Based on this work we concluded the need for spacers between the steroid moiety and the basic region peptides for subsequent designs (chapter 5).

Since the cMyc-Max was a considerably more complicated TF to mimic the project was continued by choosing the bZip Leucine Zipper as a model TF rather than the cMyc-Max. For this purpose, during the PhD of Lieselot Carrette[80], a mimic of the bZIP Leucine Zipper GCN4 TF was designed, synthesized and tested for DNA recognition. The synthesis route was based on solid phase linear coupling of the amino acids forming the basic region GCN4 onto a deoxycholic acid scaffold derivative attached to a solid support. Spacers were also incorporated between the scaffold and the first residue of each peptide chain to alleviate the rigidity imposed by the steroid nucleus for the first coupling and to give the final construct a certain

degree of flexibility, as was previously demonstrated necessary on the bHLH models[79]. In view of the expected difficulties for the linear synthesis of two such long peptide chains at the cholic acid scaffold, coupling reactions were carried out under microwave irradiation. The N-terminus was capped with 4-acetamido-benzoic acid as UV active moiety for HPLC monitoring. After cleavage from the resin by UV irradiation and deprotection of the side chains, the final construct was purified by RP-HPLC. Due to the highly complex linear nature of the synthesis, the final product could not be obtained in high purity. The product obtained after purification was subjected to an EMSA anyway to check its DNA binding affinity. A band corresponding with the DNA-peptide complex formation was obtained despite the low purity of the construct encouraging us to pursue this project. Due to the impurities of the final product, however, smearing bands were observed (figure 2.12).

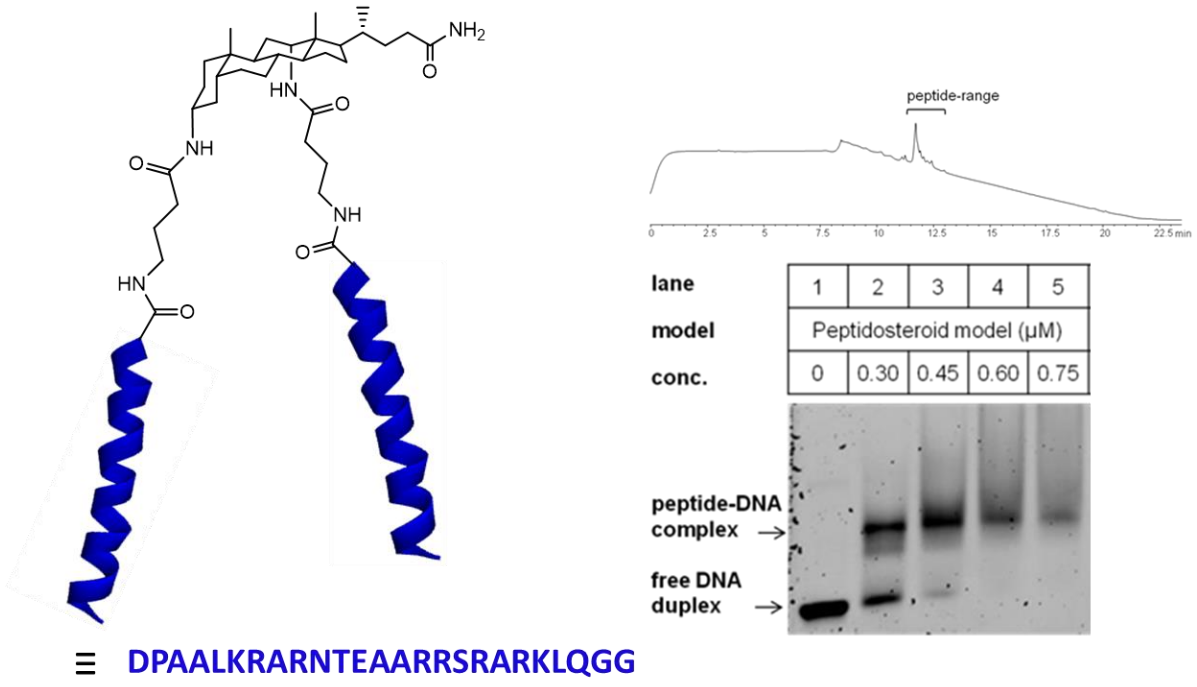


Figure 2.12: Final construct, HPLC trace (Jupiter C4 300Å in 0-100% ACN in 15 min) and EMSA for the first generation dipodal peptidosteroid model.

In conclusion, although the linear synthesis approach was successful, the low yield, extended synthesis route and poor purity of obtained compound led us to develop a convergent strategy for the incorporation of the peptide basic regions onto the bile acid-based scaffold.

2.8.2 Artificial dimerization via a steroid scaffold

As previously observed[79], spacers between the steroid nucleus and the peptides are mandatory for DNA binding recognition. These linkers should be bi-functionalized: One end to ensure the coupling to the amine functionalities of the scaffold and the other end to conjugate the peptides. Due to the versatility of click chemistry in peptide bioconjugation, we decided to explore the use of the Copper-Alkyne-Azide-

Cycloaddition (CuAAC) as the strategy for the appendage of the DNA-binding peptides to the scaffold. For this reason, one end of the spacers contains a carboxylic acid and the other end contains an alkyne or azide functionality. The geometric, steric and electronic properties of the 1,2,3-triazole resemble a trans-amide bond while also affording resistance to enzymatic degradation[47,81–83], hydrolysis and oxidation, making it an attractive heterocyclic moiety to replace unstable linkers under physiological conditions in biologically active compounds. In addition, successful replacement of two amino acids in α-helical peptides by a triazole unit has been shown to not significantly influence the secondary peptide structure[47]. To date, the triazole has only been used once to assemble an apolar tripeptide onto bile acid scaffolds[84] , despite the increasing interest of these constructs in diverse applications such as HIV inhibitors[85,86] and immunogens for vaccine development[87]. To our knowledge, long polar peptides have not been assembled on bile acid scaffolds in this manner due to the highly challenging nature of the conjugation when unprotected amino acid side chains are present. Although different synthetic conditions were applied, the planned synthetic route could not be performed.

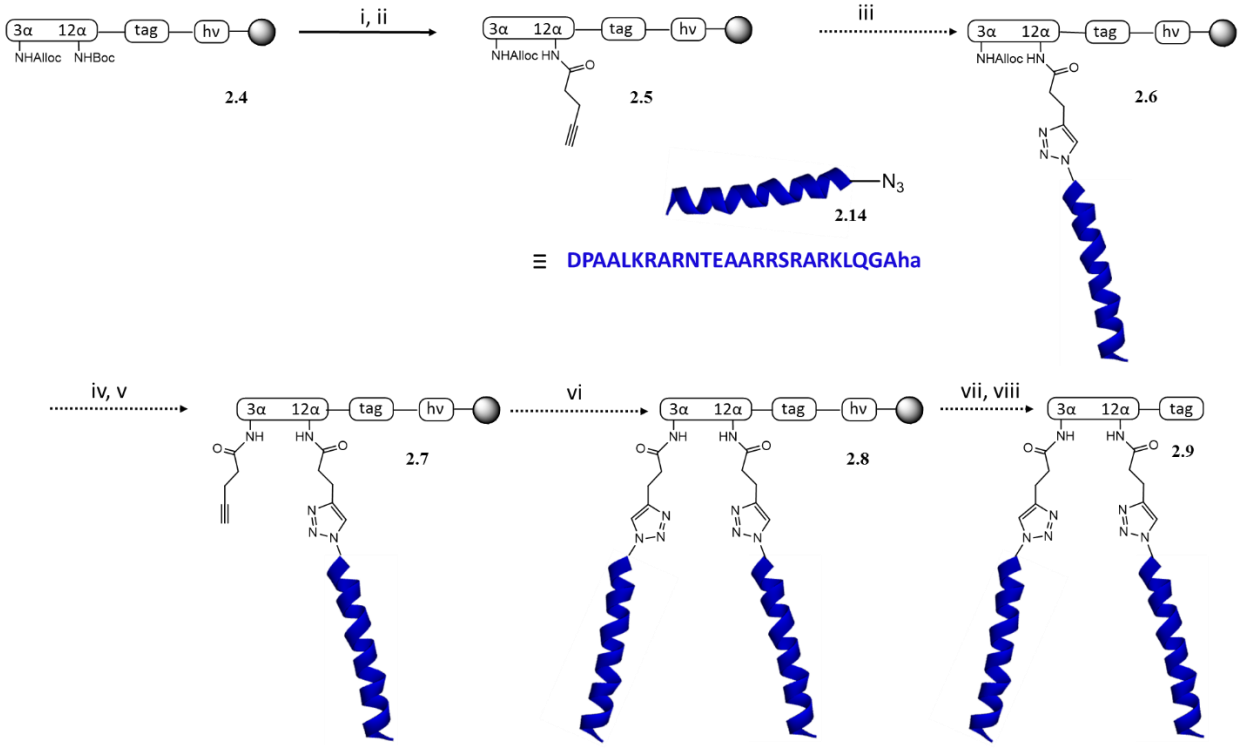


Figure 2.13: Scheme of the convergent synthesis route on solid phase towards the final dipodal peptidosteroid conjugate. The Aha term is derived from Azidohomoalanine. i) 20% TFA/DCM, 2h, rt. ii) Pentynoic acid, PyBOP, DIPEA, DMF, 4h rt. iii) Cu(I), DMF (different conditions). iv) Alloc deprotection. $Pd(PPh_3)_4$, Bu_3SnH, morpholine, DCM 2h rt. v) Pentynoic acid, PyBOP, DIPEA, DMF, 4h rt. vi) Cu(I), DMF (different conditions). vii) UV irradiation. viii) TFA/TIS/H_2O (95:2.5:2.5) 3h, rt and precipitation in cold ether.

The first step of the synthesis is the coupling of the scaffold to a solid support by using the carboxylic acid at C24, performing the coupling twice with HATU (1-[Bis(dimethylamino) methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate). The solid support is pre-equipped with the Holmes photosensitive nitroveratryl-based linker[88] and an aminomethyl benzoic acid moiety using PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) as coupling reagent. These moieties ensure a combination of sufficient reactivity and adequate UV-sensitivity for HPLC monitoring. UV irradiation as an orthogonal and mild cleavage protocol allows the semi-online monitoring of the reaction (figure 2.14).

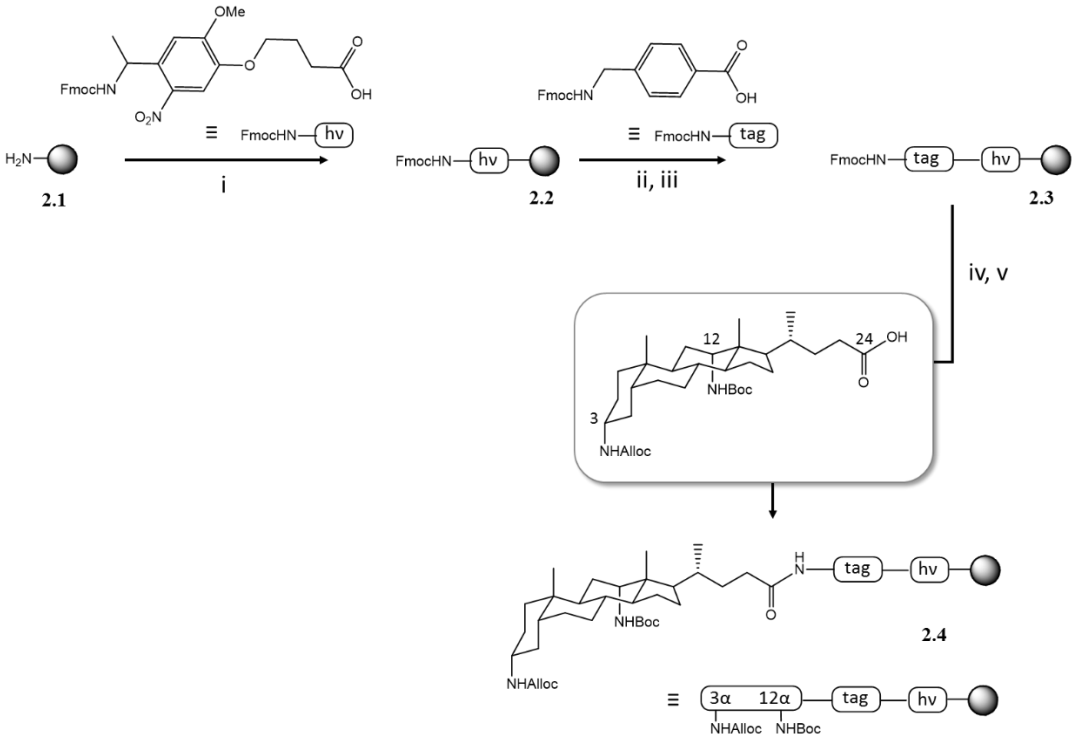


Figure 2.14: Scheme of the functionalization of the solid support with the photolabile linker, the active UV moiety and the scaffold. i) PyBOP, DIPEA, DMF, 3h, rt. ii) 20% piperidine/DMF 2 x 30 min. iii) PyBOP, DIPEA, DMF, 3h, rt. iv) 20% piperidine/DMF 2, 5, 15 min. v) HATU, DIPEA, DMF, 2 x 4h rt.

After the attachment of the scaffold to the solid support, a PyBOP mediated coupling of the linker to the C12 position is performed. Although the final construct is a homodimeric one (peptide sequences at C12 and C3 are equal), the synthesis can not be carried out in parallel due to the difference in reactivity of the amines at C3 and C12. The C12 position is more sterically hindered, and thus reactions proceed faster at C3 and increase the sterical hinderance of position C12 for further reactions. Boc deprotection (20% TFA (trifluoroacetic acid) in DCM (dichloromethane)) affords the first anchoring point for the linker, functionalized with either an alkyne or azide. Pentynoic acid and 5-azidopentanoic acid were chosen as spacers.

The basic region peptides were synthesized on 2-chlorotrityl chloride polystyrene resin. The first residue at the C-terminus bearing an alkyne or azide functionality at the side chain (Fmoc-propargylglycine-OH (Fmoc-Pra-OH) or Fmoc-azidohomoalanine-OH (Fmoc-Aha-OH)) is manually attached using PyBOP. Fmoc-Gly-OH is subsequently coupled manually to avoid epimerization at the first residue coupling. The following residues were coupled using an automated peptide synthesizer using Fmoc/tBu SPPS with HBTU (*N*,*N*,*N*′,*N*′-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate) as coupling reagent and 20-40% piperidine in NMP (N-methyl-2-pyrrolidone) for Fmoc deprotection. The N-terminus is capped with 4-acetamido-benzoic acid for the synthesis progress monitoring by UV in HPLC. Cleavage from the solid support is performed with AcOH/TFE/DCM (acetic acid/trifluoroethanol/DCM) (1:1:8) in order to keep the peptide protected during subsequent modifications at C3 (figure 2.15). After cleavage, the resin is washed with cleavage cocktail and the peptide is precipitated in cold ether (-20°C) by centrifugation at 0-4°C. The peptides were analyzed by RP-HPLC, LC-MS and/or MALDI. The peptide is not further purified and used as such for subsequent reactions.

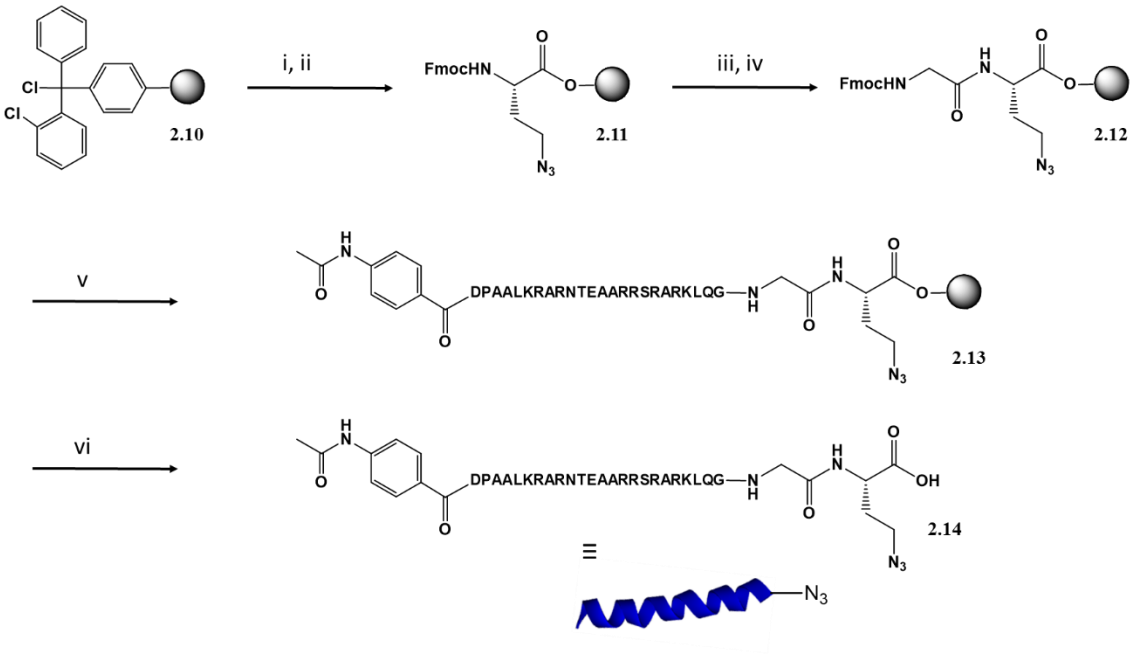


Figure 2.15: Synthesis route towards an azide-functionalized GCN4 basic region peptide. i) Fmoc-Aha-OH, DIPEA, Dry DCM, rt, 3h. ii) MeOH/DIPEA, DCM, 3 x 2 min. iii) 40% piperidine in DMF 2 x 30 min. iv) Fmoc-Gly-OH, PyBOP, DIPEA, Dry DMF, rt, 2h. v) automated synthesizer. HBTU, DIPEA, 2 x 40 min; 20% piperidine in DMF, 2, 5, 25 min. vi) AcOH/TFE/DCM (1:1:8), 4h, rt.

For the conjugation of the peptide to the solid support bearing the complementary functionalities for the CuAAC, a broad range of conditions were attempted (figure 2.16). Different solid supports, catalyst, reductants, solvent systems, bases and additives were tried. The reaction was performed at room temperature and at different temperatures ranging from 40°C to 80°C under MW irradiation. Unfortunately,

under none of these conditions, the desired peptidosteroid conjugate was obtained. The unsuccessful results can be explained by the interaction of the peptide chain with the resin and the sterical hindrance imposed by the solid support to allow the peptide to reach the C12 position.

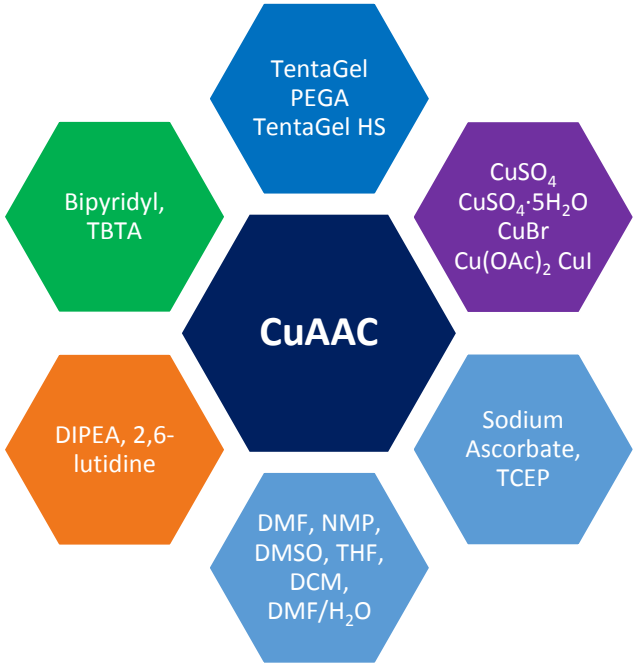


Figure 2.16: Diverse conditions used for CuAAC.

2.8.3 Rethinking the strategy: going back to the roots

After several trials to achieve the conjugation on solid support, we realized that synthesis on a solid phase was not feasible. However, as the only advantage of solid phase was to facilitate the final purification of the conjugate, we believed there was more to gain by switching to solution phase. As the presence of the resin impedes the attachment of the peptide to the scaffold, a solution phase strategy seemed a more suitable method for the conjugation of long peptides to an amphiphilic building block. Additionally, it presents several other advantages:

- ✓ The expensive orthogonally protected diamino scaffold is not required, as appendage of the peptides at both positions can be achieved in parallel and by ester bond formation. Additionally, deoxycholic acid is cheap and commercially available.
- ✓ There is no need for excess of reagents, especially the peptide, to ensure completion of the reaction.
- ✓ Synthesis could be scaled up with more ease as compared to solid phase.
- ✓ Reaction monitoring by HPLC is possible. Moreover, direct analysis of the reaction mixture can be accomplished by MALDI and RP-HPLC.

Encouraged by these benefits, we designed a new synthetic route entirely based on reactions in solution. Deoxycholic acid is commercially available and can be used directly as a scaffold for peptide dimerization of the GCN4 basic region. Indeed, the alcohols can be esterified with the linkers to provide a dialkyne or diazide scaffold, after protecting the carboxylic acid functionalities. Following this route, we designed and synthesized four different scaffolds based on deoxycholic acid as a starting material (figure 2.17). They differ amongst each other with respect to the spacers between the peptide and the steroid skeleton, which have different lengths, rigidities and functionalities. This provided a means of evaluating the factors that affect recognition of DNA in these synthetic transcription factor models. For this purpose, deoxycholic acid was modified at the alcohol positions by attachment of linkers. The linkers chosen for the study were: pentynoic acid, azido glycine, 4-azidomethyl-benzoic acid and (N-propynoylamino)-p-toluic acid (PATA). The PATA linker has been specifically developed for bioconjugation purposes as an active alkyne for preparation of peptide-oligonucleotide conjugates via CuAAC. Funtionalization of the steroid nucleus was performed by Steglich esterification, affording final scaffolds (**2.17, 2.18** & **2.19**). In case of the PATA as linker (**2.21**), the diamino derivative of deoxycholic acid proved necessary for the coupling of the linker, as esterification gave rise to byproducts due to high reactivity of the alkyne.

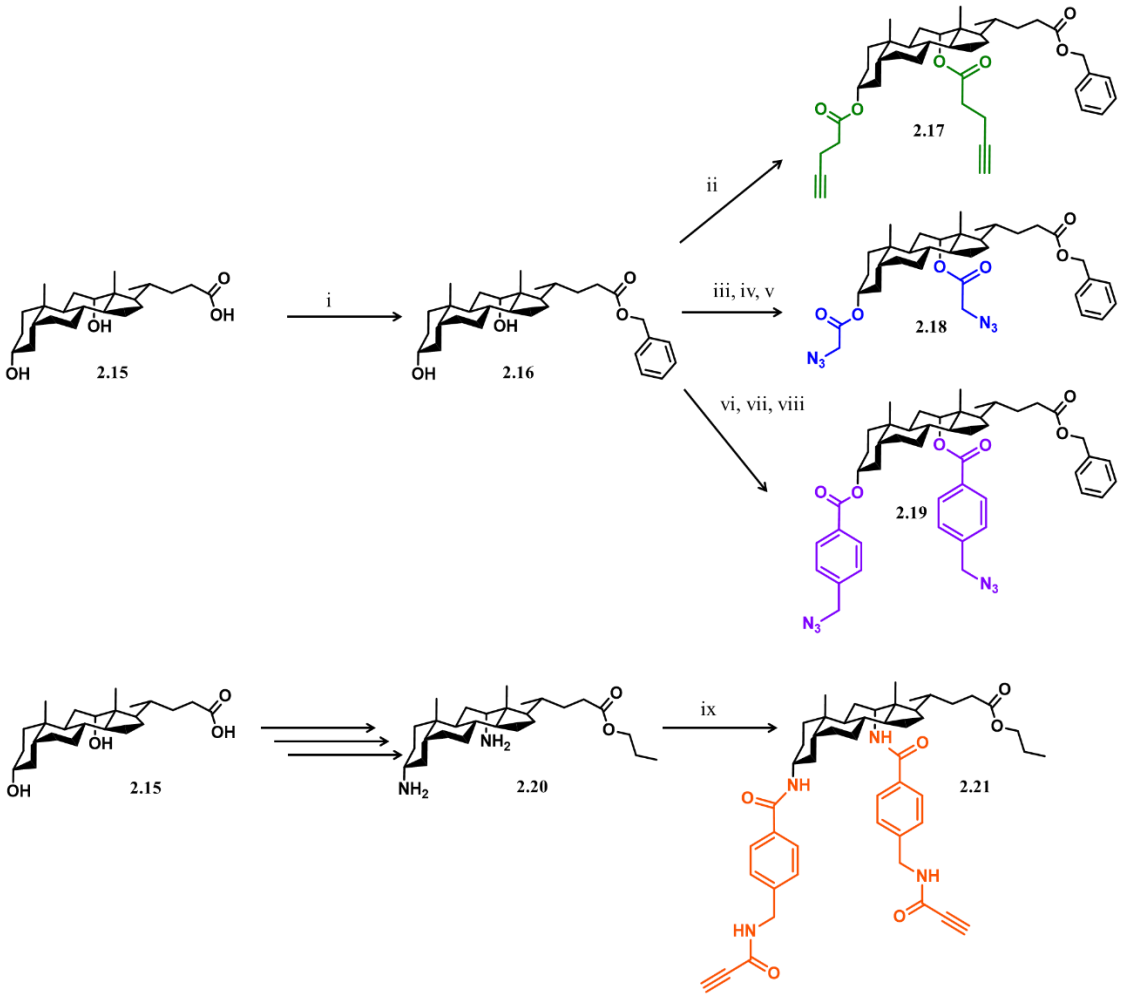


Figure 2.17: Synthetic routes towards deoxycholic acid-based scaffolds. i) DBU, BnBr, DMF 50°C, 24h. ii) Pentynoic acid, DCC, DMAP, DCM 16h, rt. iii) Boc-Gly-OH, DCC, DMAP, DCM 16h, rt. iv) 20% TFA/DCM, 2h, rt. v) K_2CO_3, $CuSO_4 \cdot 5H_2O$, Imidazole-1-sulfonyl azide·HCl, MeOH:THF (1:1) 3h, rt. vi) 4-Boc-aminomethyl benzoic acid, EDC, DMAP, DCM/THF (1:1) 16h, rt. vii) 20% TFA/DCM, 2h, rt. viii) K_2CO_3, $CuSO_4 \cdot 5H_2O$, Imidazole-1-sulfonyl azide·HCl, MeOH:THF (1:1) 3h, rt. ix) HATU, DIPEA, DMF, 16h, rt.

To obtain peptides of higher purity they were synthesized using a Rink-Amide Chemmatrix resin, which is a pure PEG resin, instead of a Polystyrene based resin (figure 2.18). The first residue at the C-terminus bearing an alkyne or an azide functionality at the side chains (Fmoc-Pra-OH or Fmoc-Aha-OH) was manually attached using HBTU. Fmoc-Gly-OH was subsequently coupled manually to avoid epimerization. The following residues were coupled in an automated synthesizer using Fmoc/tBu SPPS with HBTU as coupling reagent and 20-40% piperidine in NMP for Fmoc deprotection. The N-terminus was capped with 4-acetamido benzoic acid to allow visualization of the peptide when monitoring the click reaction by UV in HPLC. Cleavage from the solid support was performed with TFA/TIS/H_2O (trifluoroacetic acid/triisopropylsilane/water) (95:2.5:2.5), 4h. After cleavage, the resin was washed with the cleavage cocktail and peptide was precipitated in cold ether (-20°C) by centrifugation at 0-4°C. The peptides were analyzed by RP-HPLC/LC-MS and/or MALDI. The peptide was not purified further and was used as such for the click reaction.

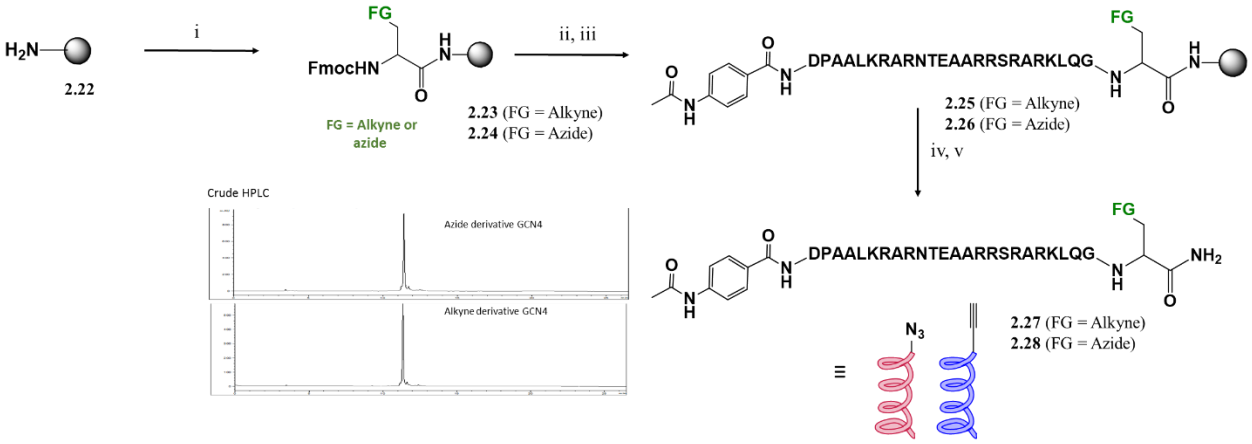


Figure 2.18: Synthesis route towards alkyne and azide-functionalized GCN4 basic region peptide and crude HPLC chromatograms (Jupiter C4 300Å, 0-100% ACN in 15 min). i) Fmoc-Pra-OH or Fmoc-Aha-OH, PyBOP, DIPEA, Dry DMF, rt, 1h. ii) 40% piperidine/DMF 2 x 30 min. iii) Automated synthesizer. HBTU, DIPEA, DMF 1h. 20% piperidine/DMF 2, 5, 15 min. iv) TFA/TIS/H_2O (95:2.5:2.5) 3h, rt. v) Precipitation in cold ether.

2.8.4 Conjugation

The GCN4 basic region peptides **2.27** and **2.28**, functionalized with an alkyne or with an azide respectively, were then attached to the central steroid core, affording four different transcription factor mimics. After a laborious optimization of the reaction conditions, the cycloaddition of the peptides to the scaffold via CuAAC was performed in a mixture of DMSO (dimethylsulfoxide)/H_2O (≈ 3:1), in order to decrease the aggregation of the peptide due to the high polarity of DMSO. Given the hydrophilic nature of the deprotected peptide and the hydrophobic nature of the scaffold, the DMSO/H_2O combination was found to be optimum for the CuAAC reaction. Moreover, it is known that polar solvents prevent aggregation of copper species. As a catalyst, $Cu(CH_3CN)_4PF_6$ gave the best results. A high excess of catalyst was needed probably due to the complexation of copper with the side chains of the peptide, containing nitrogen functionalities especially the amine of lysine and the guanidinium group of arginine. Unexpectedly, an excess of scaffold was required for the completion of the reaction, the dipodal construct still being favored over the monopodal one under these conditions, The use of reductants or additives was not necessary, commonly used to prevent oxidation of the catalyst during the reaction. The reaction was complete after 3 hours at room temperature under argon (figure 2.19). Final purification of the reaction mixture was achieved via RP-HPLC.

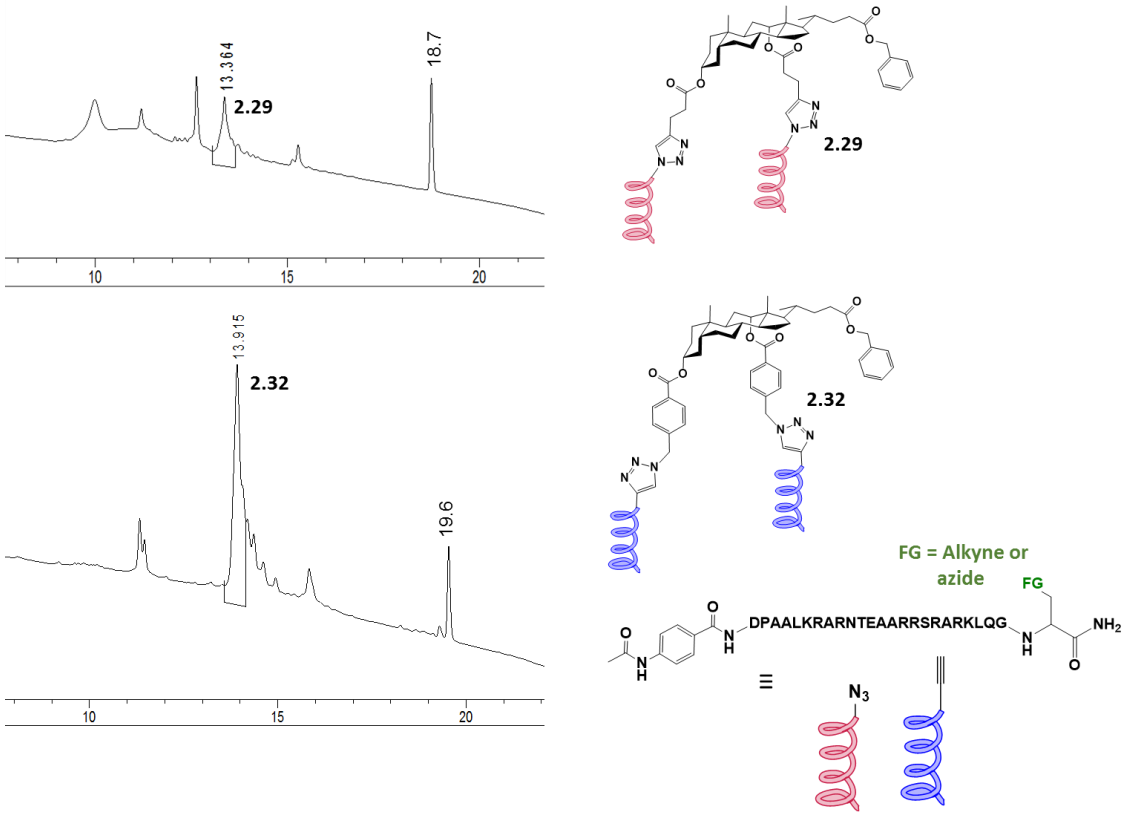


Figure 2.19: crude RP-HPLC for the reaction mixture after 3h of compounds **2.29** (top) and **2.32** (botton). Remaining scaffold **2.17** and **2.19** is visible at 18.7 and 19.6 min respectively. Final conjugates **2.29** and **2.32** are present at 13.364 and 13.915 min respectively. (Jupiter C4 300Å; 0-100% ACN in 15 min).

Apart from the above mentioned advantages of triazoles, the reaction conditions are compatible with the presence of all unprotected amino acids in the sequence. It has been reported that sodium ascorbate can lead to covalent modification when residues like lysine and arginine are present, to protein aggregation due to the formation of a dehydroascorbate byproduct and other reactive oxygen species generated by copper when sodium ascorbate is used as a reductant[89,90]. No side effects were observed in this case as no reductant is used in the conjugation. Through this work we have optimized the conditions for this conjugation which could be used in the future to synthesize constructs involving the assembly of peptides on a bile acid scaffold.

Purification of final constructs after completion of the reaction was possible via RP-HPLC (figure 2.20), affording compounds **2.29, 2.30, 2.31** and **2.32** in high purity for DNA binding studies.

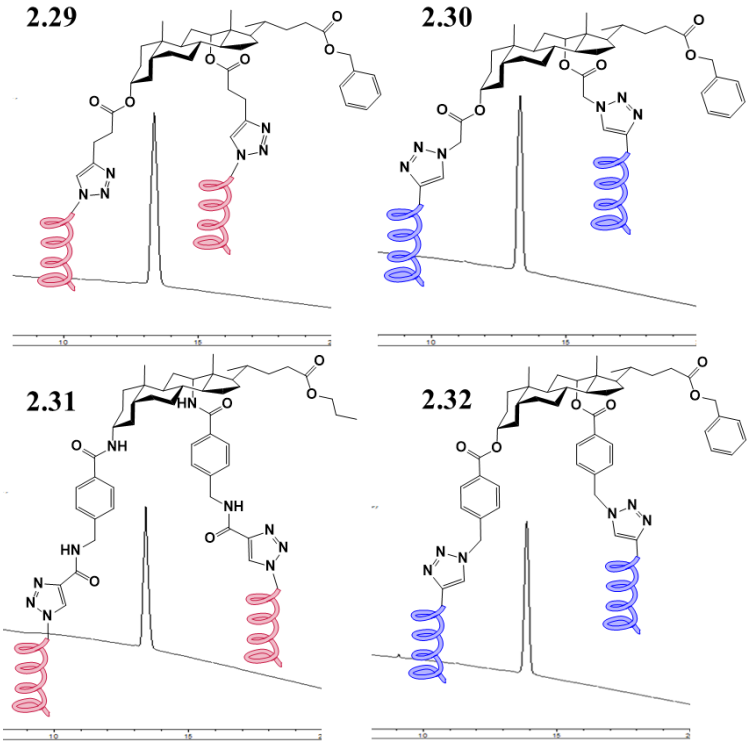


Figure 2.20: Structure of final constructs **2.29**, **2.30**, **2.31** and **2.32** and their RP-HPLC chromatograms after purification (Jupiter C4 300Å, 0-100% ACN in 15 min).

2.9 Concentration determination via ERETIC-NMR

Quantification of both DNA and peptide components is an essential step in the EMSA process. Concentration of DNA strands were provided by the commercial source. On the other hand, concentrations of final peptidosteroid conjugates **2.29, 2.30, 2.31** and **2.32** needed to be determined accurately, as minor errors on concentration determination can compromise the study of DNA binding capability. Current electrophoretic evaluation was originally attempted via weighed amounts of isolated compounds. Because of the unpredictable salt formation with trifluoroacetic acid during cleavage and deprotection steps, and obvious error of the balance when less than 1 mg was weighed, alternative strategies were considered.

Despite the broad use of UV-spectroscopy for concentration determination, it was discarded as method to calculate the concentration due to the lack of a moiety with a well-known ε-value in the presence of the benzyl ring in the scaffold. To improve reliability of present data, ^{1}H-NMR spectroscopy was exploited as analytical method for the quantification of the concentration. The ERETIC (Electronic REference To access In vivo Concentrations) method was used for determining absolute concentrations in a high-resolution 500 MHz NMR spectrometer. Introduced by Akoka et al. in 1995[91], ERETIC relies on an electronically-generated reference signal. Once calibrated against an external standard solution, determination of absolute sample concentration was straightforward via direct proportionality. Description of technical details is beyond the scope of present work. Well-resolved aromatic signals from the benzyl ring at the C24 of the scaffold, capping moieties and 5-H triazole were used to quantify final compounds. ^{1}H-NMR was performed in D_2O and the aromatic region was amplified for better integration due to the low concentration of the sample. The benzyl rings at the N-termini of the peptides generates a doublet of doublets at 7.7 ppm and 7.5 ppm, each integrating by 4 protons. The two 5-H protons from the triazoles appear in the same shift as the doublet at 7.7 ppm of the benzyl ring (figure 2.21). The correlation between the integral values and the reference sample gave the concentration of the sample: 34 μM.

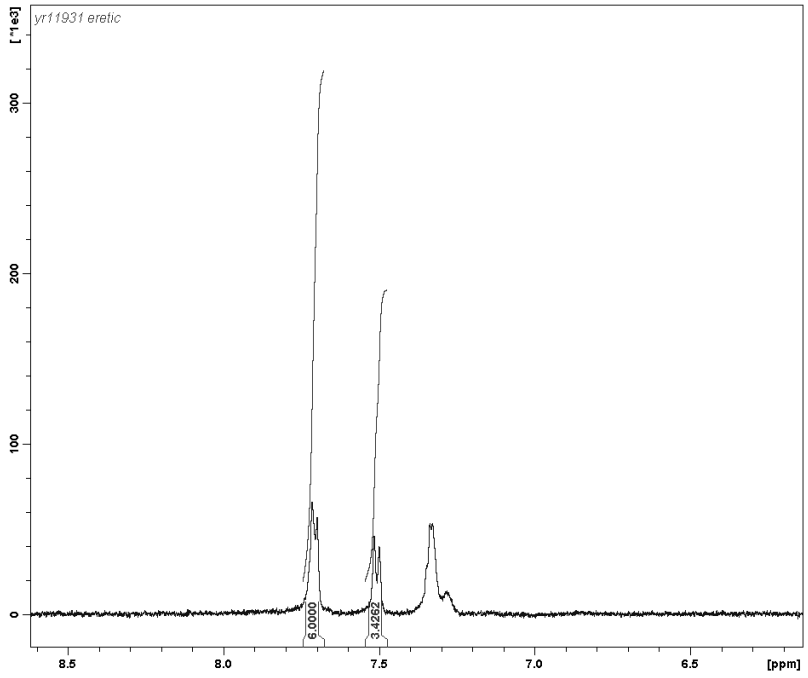


Figure 2.21. ^{1}H-NMR (D_2O, 500 MHz) of the aromatic region of compound **2.29** an integration of the signals corresponding to the 4-acetamido-benzoamide moieties at the N-termini of the peptides.

2.10. DNA Binding studies using Electrophoretic Mobility Shift Assay

2.10.1 EMSA Protocol

The protocol is described in section 10.4.

DNA binding affinities of mimics **2.29, 2.30, 2.31** and **2.32** were evaluated using an Electrophoretic Mobility Shift Assay (EMSA). Qualitative study of the interaction of mimics was based on titration of a DNA sequence containing the ATF/CREB recognition site (5' – CGG ATG ACG TCA TTT TTT TTC – 3') and Random (5' – GCG CGA GAA GGA AAG AAA GCC GG – 3') at a concentration of 0.167 μM in all the wells with increasing concentrations of compounds **2.30, 2.31** and **2.32**, lanes 1-10 contained respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 μM. The range of concentration of the mimic **2.29** in the wells were optimized for better visualization of the binding, and lanes 1-10 comprised respectively 0, 0.04, 0.12, 0.17, 0.20, 0.22, 0.26, 0.30, 0.35, 0.44 μM. As shown in figure 2.21, constructs possessing an aliphatic spacer **2.29** and **2.30** showed DNA affinity, while models with aromatic linkers **2.31** and **2.32** did not show a shifted band corresponded to the formation of peptide-DNA complex. The fact that a band of the dsDNA did not appear in the gel at the lanes in which higher concentration of peptides were added, was due to the unspecific interaction of the peptide with the DNA. Because of the electrostatic interactions that occur between the peptides and the dsDNA, the ratio m/z is increased due to an increase in the mass and a reduction in the total charge thereby resulting in their aggregation in the wells.

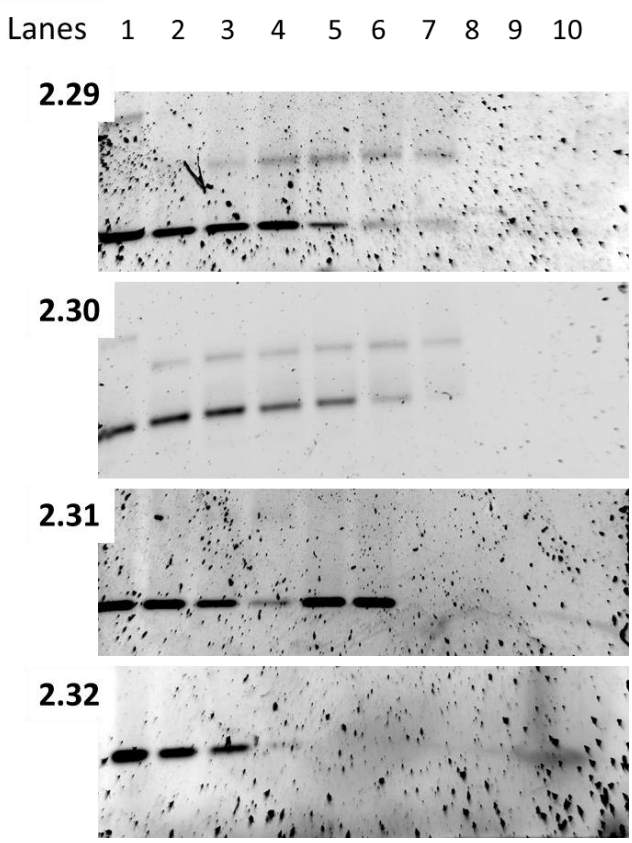


Figure 2.22: EMSA titration of the four constructs to the CRE binding site stained with SybrGold.

In order to check specificity, compounds **2.29** and **2.30** were tested with the random dsDNA sequence. Concentration of DNA in the wells was 0.167 μM and the concentration of the peptide **2.29** in lanes 1-10

were 0, 0.04, 0.12, 0.17, 0.20, 0.22, 0.26, 0.30, 0.35, 0.44 µM respectively, and for peptides **2.30**, **2.31** and **2.32** were 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM respectively (figure 2.22).

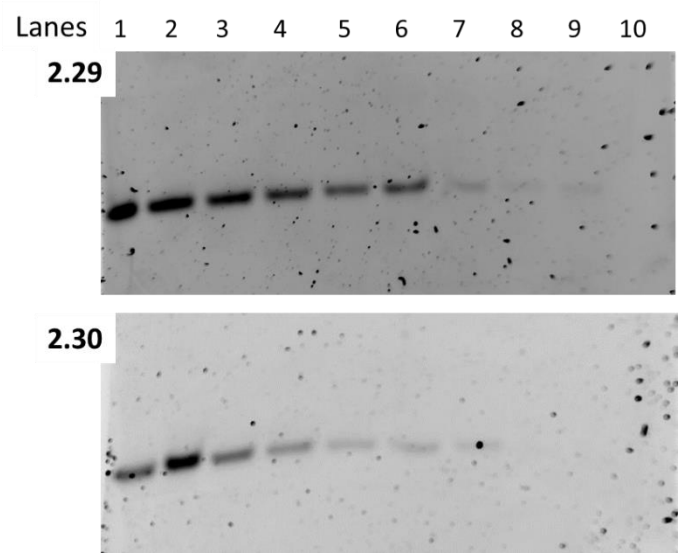


Figure 2.23: EMSA titration of binding constructs to the random sequence stained with SybrGold.

2.10.2 ^{32}P-Radiolabelled EMSA protocol

The protocol is described in section 10.5.

For the quantification of the affinity, ^{32}P-radiolabelled duplex DNA sequences containing the ATF/CREB recognition site and the random sequence were used, with increasing concentrations of compounds **2.29**, and **2.30**. From the electrophoresis experiments, we can confirm that compound **2.29** shows higher DNA affinity than **2.30**. Calculation of the Kd values is ongoing. Based on this we can conclude that aromatic rigid linkers do not allow the GCN4 basic region to adopt the correct orientation to bind the major groove and a flexible linker can favor the correct positioning of the peptides. Moreover, as compound **2.29** binds at lower concentration than compound **2.30**, a longer, more flexible linker seems to be favorable for enhanced binding.

Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water (0, 0.5, 0.625, 0.75, 0.875, 1, 1.125 and 1.25 µM).

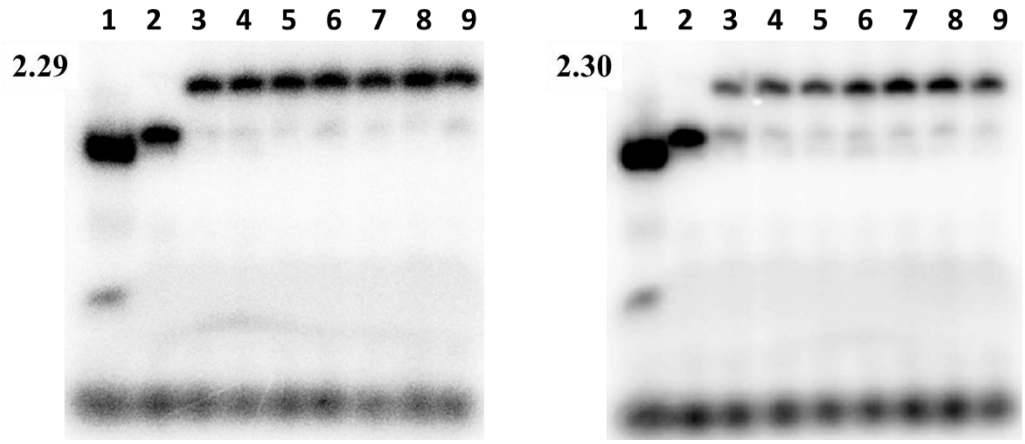


Figure 2.24. EMSA titration of the dipodal peptidosteroid conjugates **2.29** and **2.30** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 5 nM: First lane in all the gels: pyrimidine strand. Lanes 2-9 contain peptide concentrations of 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 µM for **2.29** and **2.30**.

2.11 CONCLUSION

In summary, we have synthesized major groove DNA binders based on peptidosteroid conjugates. The study of the influence of the spacer between the scaffold and the DNA recognition unit allowed us to further understand the characteristics that have an effect on DNA interaction. Our synthetic protocol was initially based on bioconjugations on solid phase. Due to the negative results following this synthetic route, we here describe the disadvantages for the attachment of peptides to a solid-supported molecule. Therefore we considered other alternatives and we here propose an optimized CuAAC protocol for the conjugation of deprotected peptides to a steroid nucleus in solution phase, allowing the application of click chemistry for the conjugation of long peptides to non-peptidic counterparts

CHAPTER 3

ARTIFICIAL DIMERIZATION VIA CYCLODEXTRIN SCAFFOLDS

Encouraged by the outcome of our peptidosteroid conjugates with regard to the synthesis and the influence of the linker on the DNA recognition ability, we envisaged the exploration of other structural features such as the distance between the C-termini of both basic regions in artificial TF models. Cyclodextrins, like cholic acid, also possess a rigid structure and therefore their dimensions are fixed and well-known. The conversion of the primary alcohols at the rim to azides allows the appendage of basic region peptides by CuAAC. By conjugation of the GCN4 basic region peptides to α, β and γ-cyclodextrin derivatives, the influence of the dimerization distance upon DNA recognition was investigated.

3.1 Structure of cyclodextrins

Cyclodextrins are natural products obtained from starch during an enzymatic degradation carried out by cyclomaltodextrin glucanotransferase and α-amylase among others. Cyclodextrins are cyclic glucose oligosaccharides. They consist of α-D-glucopyranose units linked via an α-1,4 bond, forming a lipophilic central cavity due to the skeletal carbons and the oxygens of the glucose units. They have, in addition, a hydrophilic outer surface. Cyclodextrins have the shape of a truncated cone due to the chair like conformation of the glucopyranose units. Therefore, the hydroxyl groups are oriented towards the exterior of the cone, with the secondary hydroxyl groups of the sugar residues at the wider edge of the cone and the primary hydroxyl groups at the narrower edge.

Most common cyclodextrins contain 6 to 8 units of glucose monomers in a ring. They are named as α (alpha)-cyclodextrin, β (beta)-cyclodextrin and γ (gamma)-cyclodextrin respectively (figure 3.1).

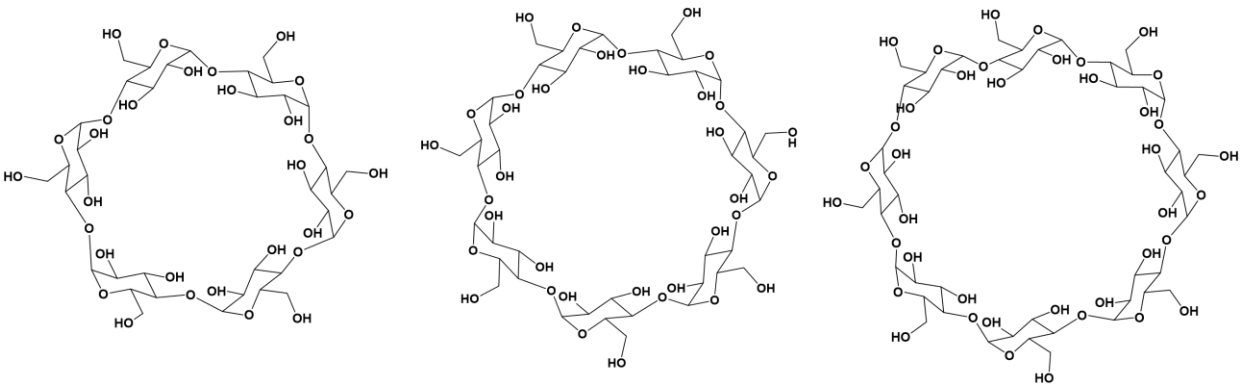


Figure 3.1: α (alpha)-cyclodextrin, β (beta)-cyclodextrin and γ (gamma)-cyclodextrin respectively in their simplest 2D form.

3.2 Applications of Cyclodextrins

The basis of cyclodextrin production was established at the end of the 19th century[92] and beginning of the 20th century[93]. The first patent on complex formation with cyclodextrins was registered in 1953[94]. Due to the high production costs, only a small scale production of cyclodextrins was possible, hampering the industrial exploitation of cyclodextrins in the pharmaceutical industry. It was not until the 70s[95,96], when advancements in synthetic technologies allowed the production of pure α, β, γ-cyclodextrins with affordable production costs. Due to the unique structure of cyclodextrins, they can form inclusion complexes with appropriately sized hydrophobic molecules. Consequently, these carbohydrates are mainly used to increase the aqueous solubility, stability, and bioavailability of drugs. But they can also be used, for example, to convert liquid drugs into microcrystalline powders, prevent drug-drug or drug-additive interactions, reduce gastrointestinal or ocular irritation, and reduce or eliminate unpleasant taste and smell of drugs by forming a complex with CDs. Recent advances in the chemistry of cyclodextrins have allowed the development of cyclodextrin conjugates with enhanced potential as drug carriers[97–101]. Nevertheless, peptide chemistry has also benefited from the properties of CDs as peptide-cyclodextrin conjugates have been previously developed for a broad range of applications, such as drug release systems[102,103], enantioselective ester hydrolysis catalysts[104], and for the creation of new types of chemosensors[105] and enzyme mimics[106]. Although the system has been intensively studied, we here show for the first time the possibility of selective dsDNA binding by peptide-cyclodextrin conjugates.

3.3 Cyclodextrin-peptide conjugates as DNA binders

In chapter 2, we explained the importance of DNA binding models to unravel the code followed by proteins which enable them to bind DNA sequence-selectively and their uses in gene therapy. We also presented the influence of the spacer between the dimerization unit and the basic region peptides on the DNA binding capacity of the final peptidosteroid conjugates. Encouraged by these results, in this chapter, we have attempted to discover other factors that influence DNA recognition in artificial TF models. Also, we further elucidate the requirements that a DNA binding model should possess to bind DNA in the correct, sequence specific manner. In this chapter, we have scanned and evaluated the importance of the distance between the

two dimerized peptides. For this purpose, we substituted the dimerization domain of the GCN4 protein by a scaffold in which the geometry could be accurately regulated. Due to the varying sizes of different natural cyclodextrins, we were able to control the distance between basic region peptides by the use of α, β and γ-cyclodextrins. Cyclodextrins were of special interest to us as we had quickly gauged the possibility of mimicking the dimerization domain of bZIP leucine zipper TF using cyclodextrins since the diameter of the primary rim almost equals the width required to append the peptides which bind to the major groove of the DNA. From a synthetic point of view, their chemical reactivity allows the selective functionalization of hydroxyl functionalities in two opposite positions at the primary rim for the dimerization.

Attachment of the basic region peptide of the GCN4 protein onto opposite positions of the primary rim of the cyclodextrins allows the control of the dimerization distance. In addition, the rigid structure of cyclodextrins and the orientation of the functionalities of the primary rim ensures the correct position of the basic region peptides and facilitated the anchoring on the major groove. Moreover, it has been observed that cyclodextrins improved the bioavailability of the attached peptides[107].

Based on the diameter of the primary rim of the α, β and γ–cyclodextrins which are 5.7, 7.8 and 9.5 Å respectively[108] (figure 3.2), three different peptide-cyclodextrin conjugates were synthesized, keeping in mind that the width of the major groove is 11.2 Å.

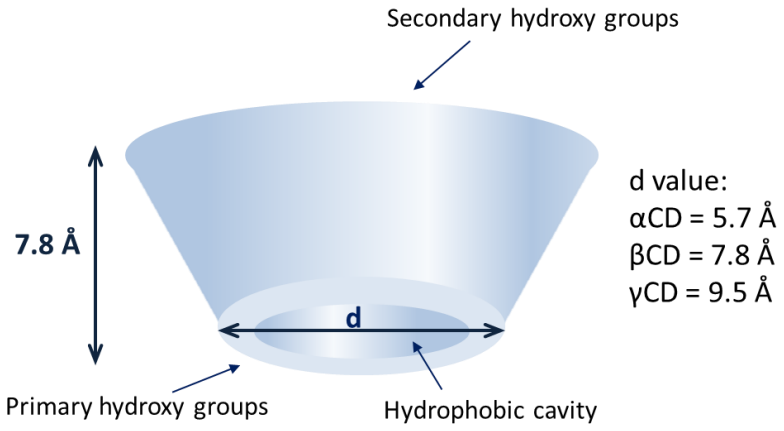


Figure 3.2: Characteristics of the cyclodextrin structure.

3.3.1 Synthesis of the cyclodextrin derivatives

Before the use of our well established click chemistry, the CDs need to be functionalized by means of chemical modifications. One of the most useful functionalizations is the introduction of an azide group in the "cyclostructure" by substitution reactions of the hydroxyl groups. Hence, the obtained azido-CDs are good candidates for our click chemistry and for attaching the GCN4 basic region peptides. In view of the expertise of the group of Tomas Kraus at the Academy of Sciences of the Czech Republic, Prague, on the selective modification of cyclodextrins, we established a collaboration to gain access to the required modified CDs as starting materials. Starting from $6^{I},6^{IV}$-dideoxy-$6^{I},6^{IV}$-dibromo-cyclodextrins[109], direct nucleophilic substitution with NaN_3 afforded $6^{I},6^{IV}$-Dideoxy-$6^{I},6^{IV}$-diazido-α-cyclodextrin, $6^{I},6^{IV}$-Dideoxy-

$6^{I},6^{IV}$-diazido-β-cyclodextrin and $6^{I},6^{V}$-Dideoxy-$6^{I},6^{V}$-diazido-γ-cyclodextrin in 90%, 83% and 70% yield respectively (figure 3.3). These compounds were kindly provided to us for further derivatisation.

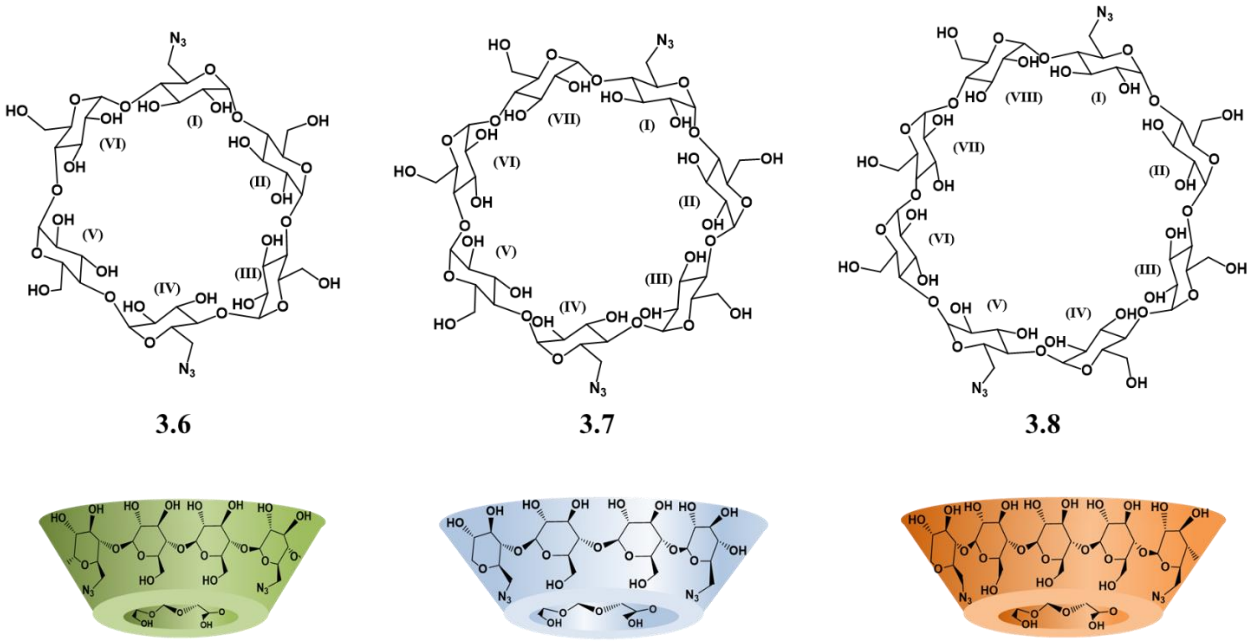


Figure 3.3: Chemical structure of diazido-cyclodextrin derivatives **3.6, 3.7** and **3.8** and 3D representation of $6^{I},6^{IV}$-diazido-α-cyclodextrin, $6^{I},6^{IV}$-diazido-β-cyclodextrin and $6^{I},6^{V}$ - diazido-γ-cyclodextrin.

3.3.2 Synthesis of basic region peptides

The GCN4 DNA recognition peptides D226-Q248 were synthesized on solid phase. Modifications of the peptide sequence were performed at the C-terminus. An alkyne was incorporated for the CuAAC conjugation to cyclodextrin. In addition, as we have previously shown in chapter 2, correct length and flexibility of linkers are a must regarding DNA recognition in artificial TF models. Through molecular modelling we inferred that direct conjugation of the basic region peptides to the cyclodextrins would not afford the desired conjugates. Direct conjugation would instead result in the peptide not adopting the proper conformation for DNA binding in the major groove due to a lack of flexibility. Thus, incorporation of a spacer between cyclodextrin-based dimerizer and peptide was once again mandatory for optimal sequence-selective DNA binding. The length and flexibility that pentynoic acid afforded to the previous peptidosteroid conjugate **2.29** was mimicked in the peptide-CD conjugates by the incorporation of γ-aminobutyric acid as a second residue at the C-terminus of the peptide. In order to check the purity of the peptide after cleavage and deprotection and to follow the reaction with the cyclodextrin by RP-HPLC, the N-terminus was capped with 4-acetamido-benzoic acid as UV absorption moiety (figure 3.4).

Figure 3.4: Modification of the GCN4 peptide for suitable attachment to CDs.

Rink-amide ChemMatrix resin was chosen as solid-support for peptide synthesis. Couplings were performed using HBTU and DIPEA. Final deprotection and cleavage was achieved with 95% TFA, 2.5% H_2O and 2.5% TIS. Precipitation in cold Et_2O and analysis on MALDI and RP-HPLC showed that the desired peptide was obtained with sufficient purity for further conjugations.

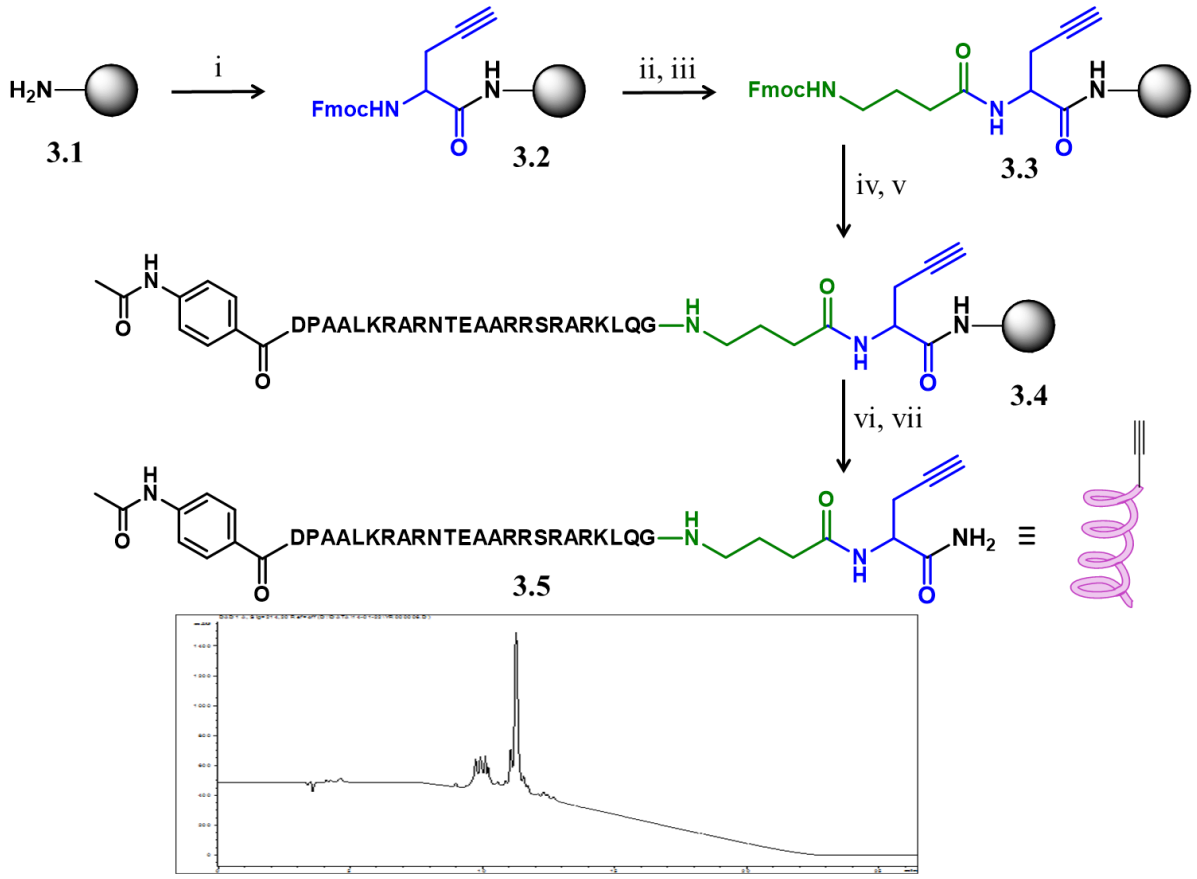


Figure 3.5: Synthesis scheme and HPLC trace of the crude modified GCN4 peptide **3.5** for conjugation to the CDs (Jupiter C4 300Å, 0-100% ACN in 15 min). i) Fmoc-Pra-OH, HBTU, DIPEA, Dry DMF, rt, 1h. ii) 40% piperidine/DMF 2 x 30 min. iii) Fmoc-GABA-OH, HBTU, DIPEA, Dry DMF, rt, 1h. iv) 20% piperidine/DMF 2, 5, 15 min. v) Automated synthesizer. HBTU, DIPEA, DMF 1h. 20% piperidine/DMF 2, 5, 15 min. vi) TFA/TIS/H_2O (95:2.5:2.5) 3h, rt. vii) Precipitation in cold ether.

3.3.3 Click with Cyclodextrins

Before the development of click-chemistry, different strategies for the formation of a covalent link between various ligands and cyclodextrins have been reported. Glycosilation, formation of thiourea or amide bridges, amidation and nucleophilic substitutions were investigated for the formation of cyclodextrin-based conjugates[110,111]. More recently, cyclodextrin conjugation via CuAAC has been investigated in depth since its discovery by the Meldal and Sharpless research groups[82,112] in 2001. Translation of optimized protocols and conditions from organic chemistry to supramolecular chemistry allowed the synthesis of a large variety of cyclodextrin-based conjugates, and therefore CuAAC has been documented as a high yielding and kinetically fast reaction for the preparation of carbohydrate conjugates. CuAAC with cyclodextrins was first introduced for biological applications in 2007 for the synthesis of different glycoconjugates as synthetic activators in cell adhesion, stimulation of monocytes and for the elaboration of a controlled drug delivery system in the oral cavity[113,114]. Other groups have applied CuAAC for the decoration of cyclodextrin with polymers[115] and for the synthesis of nontoxic drug delivery carrier systems based on stable helical polypeptide-grafted cyclodextrin bioconjugates[116].

We report here for the first time the application of cyclodextrin conjugates for sequence-specific DNA recognition and the optimized conditions for CuAAC of cyclodextrin derivatives with relatively long, deprotected peptides (figure 3.6). Inspired by the properties of 1,2,3-triazoles in resembling a trans-amide bond and affording resistance to enzymatic degradation[47,81–83], hydrolysis and oxidation, it made an attractive heterocyclic moiety to replace unstable linkers under physiological conditions in biologically active compounds. In addition, successful replacement of two amino acids in α-helical peptides by a triazole unit has been shown to not significantly influence the secondary peptide structure[47]. Moreover, through this work we have shown that our conditions can be easily translated to the synthesis of other peptide conjugates as the reaction protocol and conditions are similar to the ones reported in chapter 2 for the functionalization of steroid-based scaffolds with peptides. DMSO/H_2O (≈ 3:1) was used as solvent in the CuAAC of the cyclodextrin-based scaffold and the peptide. DMSO was required to dissolve the scaffold, while the peptide was quite soluble in water. Moreover, the DMSO/water mixture provided a polar medium to avoid the aggregation of the peptide and copper species. As a catalyst, $Cu(CH_3CN)_4PF_6$ gave the best results as was observed for the peptidosteroid conjugate synthesis. Unfortunately, once again a high excess of catalyst was needed due to the complexation of copper with the side chains of the peptide, containing nitrogen functionalities especially the amine of lysine and the guanidinium group of arginine. As it was previously observed, an excess of cyclodextrin derivative was required for the completion of the reaction, and the dimer formation was favored over the monomeric one. The use of reductants or additives was not necessary. The reaction was stopped after 3 hours at room temperature under argon. Final purification of the reaction mixture was achieved via RP-HPLC (figure 3.7). As mentioned in chapter 2, the absence of sodium ascorbate as reductant in the reaction mixture avoided side reactions already observed in peptides containing lysines and arginines [89,90].

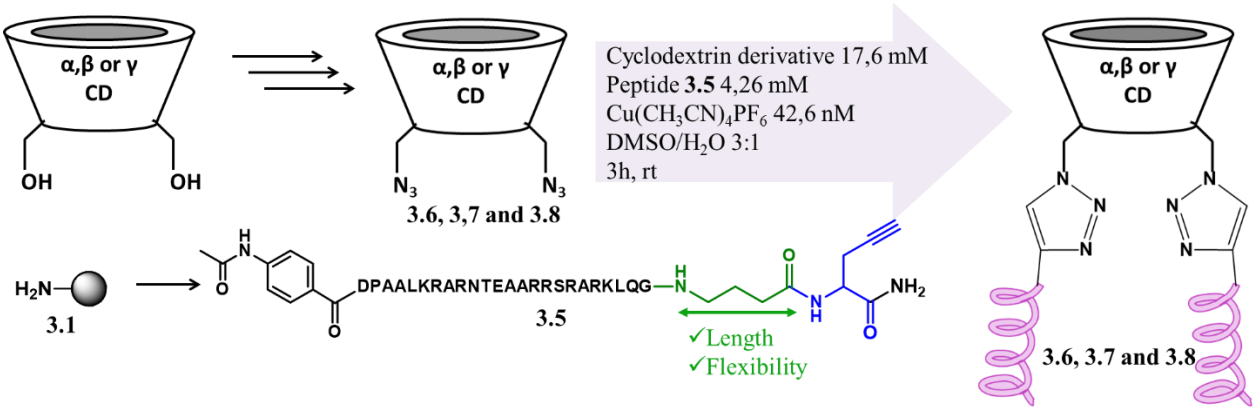


Figure 3.6: Click reaction with CD.

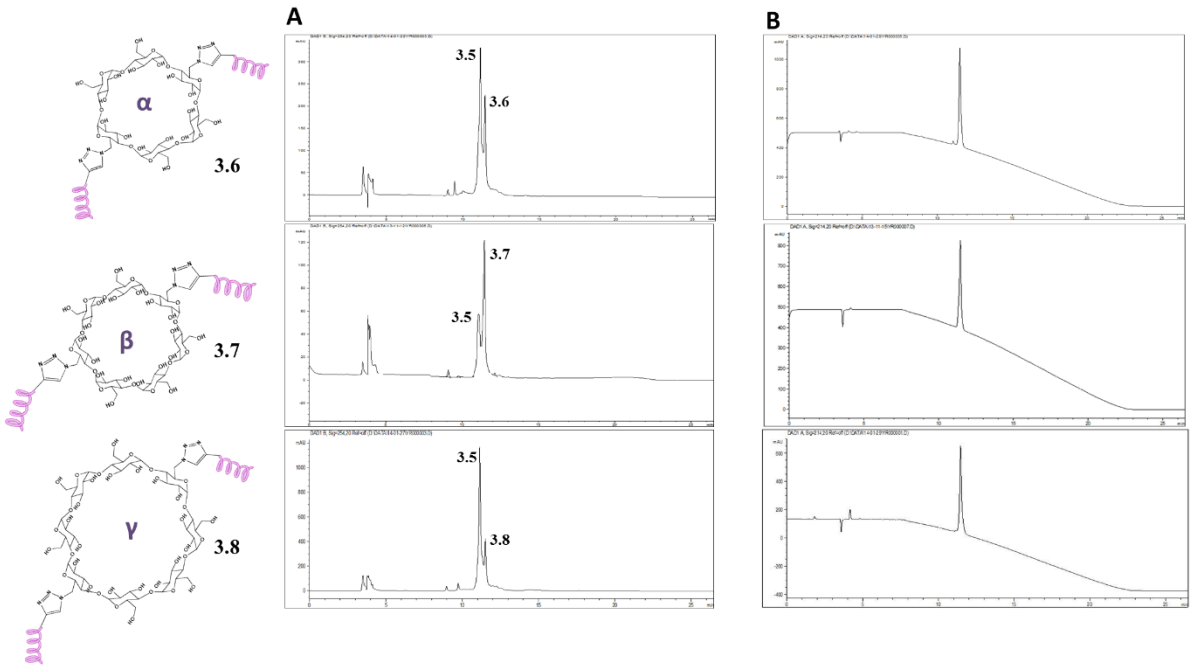


Figure 3.7: HPLC traces of crude (A) and final compound **3.6**, **3.7** and **3.8** (B) (Jupiter C4 300Å, 0-100% ACN in 15 min).

After the synthesis and purification the cyclodextrin derivatives **3.6**, **3.7** and **3.8** functionalized with GCN4 basic region peptides and the dimers were subjected to DNA binding studies.

3.4 Concentration determination via ERETIC-NMR

As reported in chapter 2, ^{1}H-ERETIC technique was chosen as method to accurately determine the concentration of the samples **3.6**, **3.7** and **3.8** for the EMSA study. Aromatic signals from the 4-acetamido-benzamide moieties at the N-termini and the 5-H of the triazoles were used as signals to estimate the concentration of the sample from the reference sample. ^{1}H-NMR was performed in D_2O and the aromatic

region was amplified for better integration due to the low concentration of the sample. The benzyl rings at the N-termini of the peptides generate a doublet of doublets, each integrating for 4 protons (figure 3.8). The correlation between the integral values and the reference sample gave the concentration of the sample: 77 µM. The two protons from the triazole were not well resolved and therefore they were not taken into account for the calculations. The same procedure was repeated for both **3.6** and **3.8** compounds, as they present the same NMR spectra at the aromatic region.

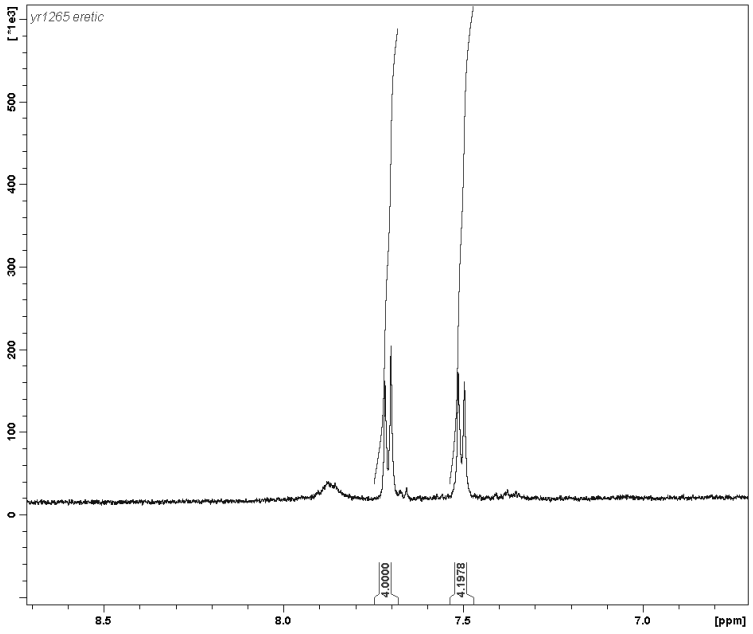


Figure 3.8. ^{1}H-NMR (D_2O, 500 MHz) of the aromatic region of compound **3.7** an integration of the signals corresponding to the 4-acetamido-benzoamide moieties at the N-termini of the peptides.

3.5 DNA binding studies using EMSA

The DNA binding affinities of mimics **3.6, 3.7** and **3.8** were evaluated using an Electrophoretic Mobility Shift Assay (EMSA). Tests were first performed in UGent using non-radiolabelled duplex DNA (figure 3.9) followed by tests in Karolinska Institute in Sweden using ^{32}P-radiolabelled duplex DNA (figure 3.10 top).

The protocol is described in section 10.4.

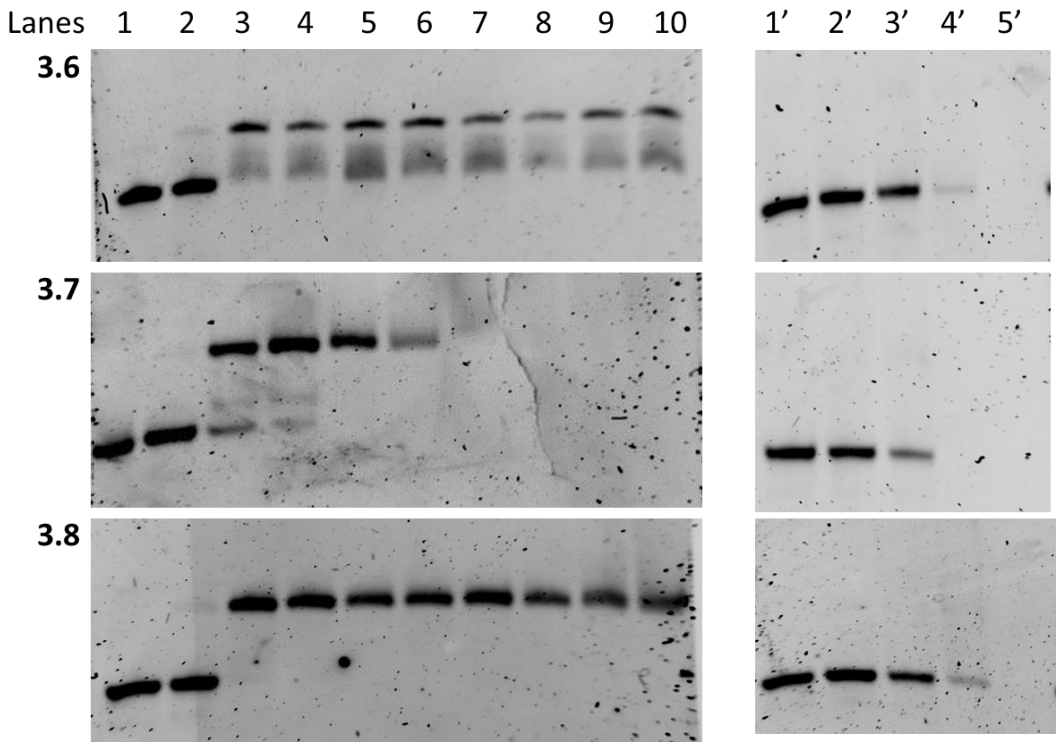


Figure 3.9: EMSA titration of the dipodal peptidocyclodextrin conjugates **3.6, 3.7** and **3.8** from top to botton to the CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3'). Lanes 1-10: dsDNA (0.167 μM); lanes 1-10: 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 μM of peptide, and to the Random sequence (5' – GCG CGA GAA GGA AAG AAA GCC GG – 3'); lanes 1'-5': 0. 0.167, 0.501, 0.668, 1.67 μM of peptide.

The study was based on the titration of a ^{32}P-radiolabelled duplex DNA sequence containing the ATF/CREB recognition site (5' – CGG ATG ACG TCA TTT TTT TTC – 3') and Random (5' – GCG CGA GAA GGA AAG AAA GCC GG – 3') with increasing concentrations of compounds **3.6**, **3.7** and **3.8**. From the electrophoresis experiments, we could calculate K_D values of 50 ± 20, 30 ± 20 and 100 ± 60 nM respectively, with compound **3.7** showing highest DNA affinity. Unlike in the case of the peptidosteroid conjugates, in which we saw that the linker had the potential to completely obstruct sequence-selective DNA recognition, the modification of distance between the C-termini did not hinder DNA binding to that great an extent, as a band corresponding to the DNA-peptide complex was visible in all the gels of cyclodextrin-peptide conjugates. We consider that it only influences the way in which the peptides recognize the DNA and the orientation of the side chains of the amino acids towards the nitrogen bases. Despite this, the perturbation was not so great as to completely mitigate DNA binding. Previous models developed by Morii and Mascareñas possess the same length of the basic region peptides (D226-Q248) and therefore are comparable with our systems. The group of Morii developed a system in which the dimerization unit was a complex of cyclodextrin and adamantane. Both counterparts were anchored to a GCN4 basic region peptide and bind only when dimerized. EMSA studies were performed with < 1 nM radiolabelled CRE, obtaining a Kd < 150 nM[56]. More recent models designed by the group of Mascareñas employed a diazobenzene moiety for dimerization, with a Kd < 5 nM using < 1 nM radiolabelled CRE.

They also reported the affinity of a disulfide-bridged GCN4 basic region, with a Kd of < 150 nM approximately for a DNA concentration of 50 nM[117]. The latest model consisted on a dual mimic in which dimerization arrangement and DNA selectivity was controlled by selected external parameters[118]. The basic region peptides were functionalized with a cysteine and a terpyridine moiety attached to the residues at the N- and C- termini of the peptide chain to achieve selective dimerization at both sites. Depending on the conditions, the disulfide-based or the metal-terpyridine complex-based dimers were preferred, and different DNA sequences could be targeted. By dimerizing the peptide in the presence of Ni^{2+}, the construct was bound to the CRE sequence with a Kd of 299±26 nM, using a DNA concentration of 100 nM, ~100 pM labeled with ^{32}P. For these 3 last models, it should be noted that the nonionic detergent (NP-40) and BSA were used during EMSA, known to decrease aggregation of peptides and proteins and therefore favor the interaction with the DNA[119,120]. A rough comparison thus indicates that our new constructs, apart from being synthesized in a straightforward and convergent manner, display comparable binding properties to those described in literature. The obtained values are further within the same order of magnitude as those calculated for the binding of bacterially expressed GCN4 and synthetic versions thereof. For instance, a dimer comprising of a 56mer GCN4 basic region (residues 226-281) also binds CRE in the nanomolar range (Kd ~ 12 nM)[121–123]. No data for the dissociation constant of the natural protein have been reported so far.

In order to better explain the results obtained from the binding pattern on the gels and the determination of the Kd values, we performed a molecular visualization of the dimerization interface of the bZIP GCN4 TF obtained by discarding the leucine zipper domain from the crystal structure of the natural protein (figure 3.10, D). The distance between the C-termini of the basic regions is shown in figure 3.10 at different perspectives of the protein-DNA interface, and is found to be 15.052 Å. The Kd values showed that derivative **3.7** was the one with the best binding capability to the CRE sequence. Therefore, we consider that beta CD is the dimerizer which allows optimal anchoring of the peptides on the major groove of the DNA. In the protein the C-termini are placed at a distance of approximately 15 Å. In case of our constructs, although the diameter of the primary rim of the cyclodextrins is known (figure 3.10, left), due to the flexibility of the linker, the exact distance between the C-termini of the peptides on **3.5**, **3.6** and **3.7** could not be predicted accurately. However, we observed that increasing or decreasing the distance between the attachment points of the peptides by the use of different CD scaffolds results in a deviation of the optimum geometry of the system. This is reflected in terms of decreased binding affinity and lower Kds. It should also be noted that these results are specific for the basic region GCN4 peptide sequence and the given spacers. In case these parameters are varied, different conclusions are expected.

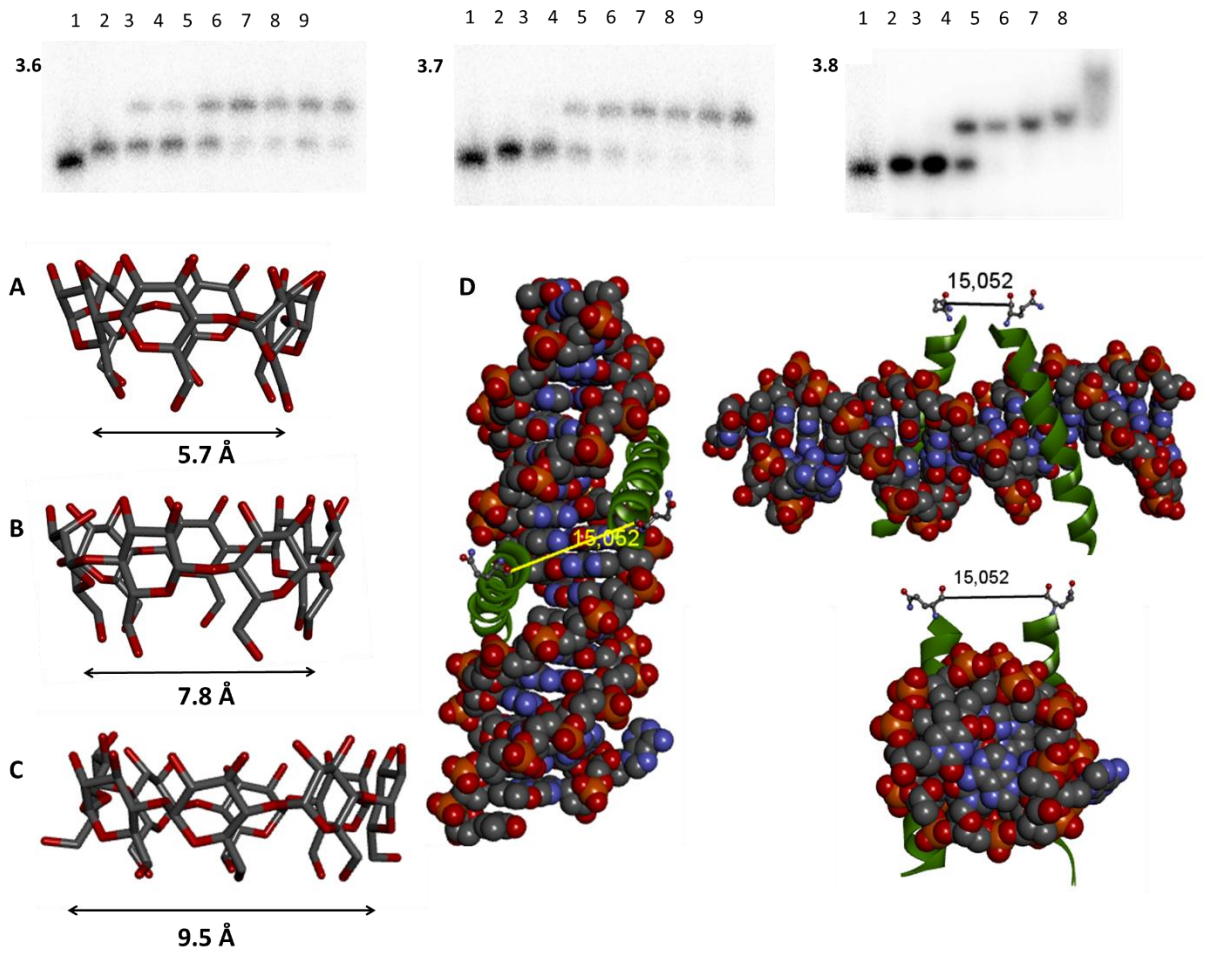


Figure 3.10: EMSA titration of the dipodal peptidocyclodextrin conjugates **3.6, 3.7** and **3.8** from left to right (top) to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 5nM: First lane in all the gels: pyrimidine strand. Lanes 2-9 contain peptide concentrations of 0, 0.5, 0.625, 0.75, 0.875, 1, 1.125 and 1.25 μM for **3.6** and **3.7**, and lanes 2-8 with 0, 0.6, 1.25, 2.5, 5, 10, 20 μM for **3.8**. Bottom: 3D structure of α, β and γ-cyclodextrin and primary rim diameter distance (A, B and C respectively) and crystal structure of the GCN4 basic region peptides appended to the major groove of the DNA from different perspectives. Distance between C-termini indicated in angstroms (PDB: 1YSA) (D).

3.6 Conclusion

To conclude, we have presented the design and synthesis of three peptide-cyclodextrin conjugates for sequence-selective DNA recognition. This was achieved by the use of α, β and γ-cyclodextrin derivatives as scaffolds for the appendage of the peptides by CuAAC. $6^{I},6^{IV}$-diazido-α-cyclodextrin, $6^{I},6^{IV}$-diazido-β-cyclodextrin and $6^{I},6^{V}$-diazido-γ-cyclodextrin were successfully synthesized and fully characterized by the group of Tomas Kraus at the Academy of Sciences in the Czech Republic. Though examples exist of

CuAAC reactions with CD derivatives[116,124–126], to the best of our knowledge long, deprotected peptides have so far not been conjugated to cyclodextrins via CuAAC. We here present optimized conditions for the anchoring of peptides to cyclodextrin units via click chemistry.

Our results indicate the usefulness of an optimized dimerization configuration between both peptides in artificial TF models. Indeed, it was shown that the distance between the anchoring points has a notable influence on DNA binding. Successful models can be obtained by trying to approach the exact features of the protein at the interface between the dimerization and the basic region domains of the bZIP TF to achieve DNA binding comparable to that of the natural TF.

CHAPTER 4

CELL PENETRATION STUDIES

*The next step after the achievement of sequence-selective dsDNA recognition was to establish "proof of concept" that this system could also be adapted for cells. With this in mind we tested the cell uptake of the peptidosteroid conjugate **2.29** and peptidocyclodextrin conjugate **3.7**, as future applications in vivo will necessitate the constructs to reach the nucleus and to bind DNA. This study was developed during the master thesis of Dorien Van Lysebetten, under the guidance of Abhishek Iyer. Benoit Louage and Prof. Bruno De Geest are thanked for their valuable contributions to this work.*

4.1 Introduction

The idea that bile acids can increase cellular uptake of peptides is proposed in many papers[66,67,69]. However long peptides conjugated to bile acids have not yet been tested for cell uptake because of the challenging conjugation between cationic, hydrophilic CPPs and a hydrophobic steroid moiety. Through this work the synthetic challenges have been reduced to great extent. The development of sequence-specific DNA binders based on peptidosteroid conjugates is directed towards new strategies in cancer therapy. Consequently, an important requirement would be that the compound should penetrate through the membrane of the cell and reach the nucleus. This in itself requires overcoming several challenges. Firstly, the peptide-steroid conjugate must be cell penetrating. Moreover, it must not be trapped in the endosomes and must be able to perform endosomal escape. Finally, it must be able to enter the nucleus. Several sequence specific DNA binders exist[55–58,127]. However, none, to the best of our knowledge, have been tested or have known to show enhanced cell penetration.

Cell penetrating peptides (CPP) define a broad class of peptides generally 5-30 residues in length which are taken up by different cell types at low concentrations while not inducing cytotoxicity[128,129] . The basic DNA binding region of the GCN4 protein, a 23-mer peptide with 6 arginine residues mainly responsible for the positive charges, is classified as a cationic CPP. Although it does not bind DNA in its monomeric form, its dimerized version is known to show high affinity towards dsDNA. In this study the dimerization has been performed via a steroid scaffold allowing DNA recognition as shown in chapter 2. In this section we have explored the cell penetration ability of the dimers **4.4** and **4.5** and the monopodal version **4.3**.

Research into the cell permeability of the peptides was carried out with the help of flow cytometry and confocal microscopy whereby the latter gives qualitative information about the localization into the cell[130]. These methods are based on measuring emitted light of a fluorophore which demanded the (re)synthesis of

all peptides described in chapter 2 and 3 with the incorporation of a fluorescent group. This was done in a fairly straightforward manner by coupling the N-hydroxy-succinimide (NHS) ester of fluorescein (figure 4.1) to the N-terminus of the peptides following solid phase synthesis. Fluorescein-labelled basic region GCN4 (**4.1**) and basic region GCN4 dimer (**4.2**) obtained as a side product in a parallel project focusing on Cysteine derivatives of GCN4 for peptide stapling, were used as controls as they do not bind DNA. Fluorescein-labelled monopodal peptidosteroid conjugate (**4.3**), although non-DNA binding was designed specifically to study of the influence of a steroid in the cell uptake of the GCN4 peptide. Fluorescein-labelled dipodal peptidosteroid (**4.4**) and peptide β-cyclodextrin (**4.5**) conjugates were also tested as they were found to possess the best DNA binding properties as shown in chapters 2 and 3.

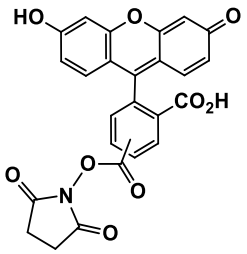
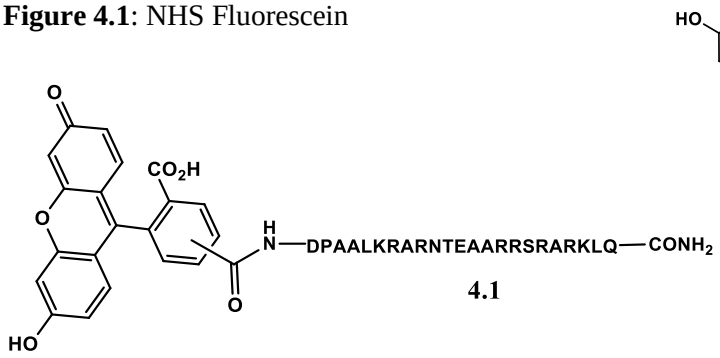
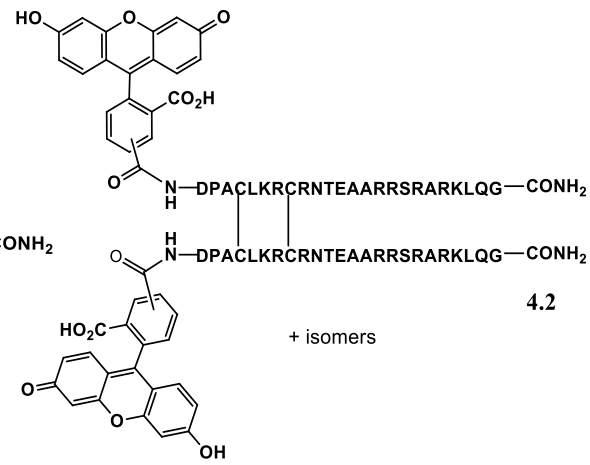


Figure 4.1: NHS Fluorescein

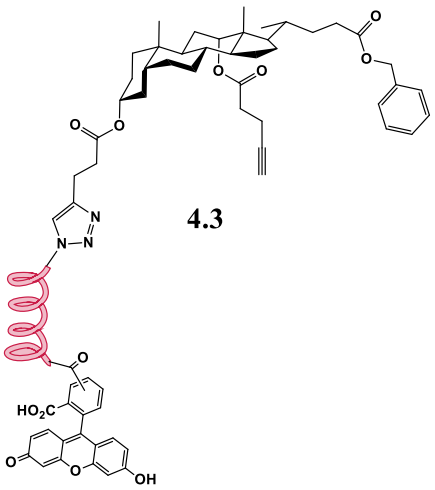


4.1

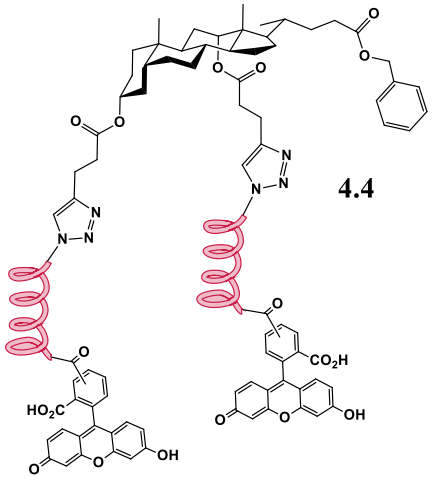


4.2

+ isomers



4.3



4.4

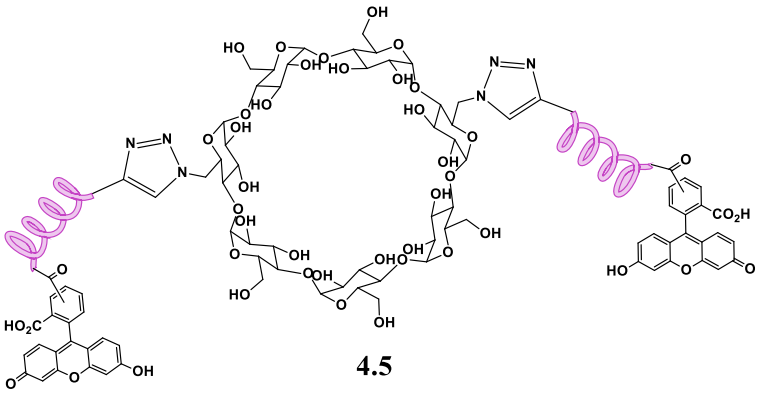


Figure 4.2: Structures of constructs synthesized for this study.

Questions about protease stability and toxicity were answered by testing the stability in serum and performing an MTT assay respectively. The specific choice of the steroid scaffold is based on its ability to enhance proteolytic stability of attached peptides[131,132], by its amphiphilicity[70], by its conformational properties ensuring correct positioning of the two appended chains[71]and by its potential to increase cellular uptake and bioavailability[68]. Additional tests were also performed to investigate the kinetics of uptake which is measured with the help of flow cytometry. Comparing the uptake at different temperatures will increase the understanding of the uptake mechanism of the peptides[133].

4.2 Serum protein complexation

Bovine serum albumine (BSA) is a globular protein consisting of 583 amino acids counting for a molecular weight around 66.5 kDa. It is the most abundant protein in fetal bovine serum (FBS) and therefore an important factor in uptake experiments. This protein has a relatively high hydrophobic character and can easily interact with other hydrophobic molecules. The GCN4 peptide could therefore stick to this abundant BSA which could as result in a decrease of the free peptide concentration drastically. In addition to the above, the bigger BSA peptide complex can be recognized by the cells as a particle which is internalized by a different mechanism. The specific type of cells determines how easily particles are internalized. For example, the RAW264.7 cells used in current thesis can easily take up particles. Therefore, it was important to check for complexation of peptides with serum proteins in order to measure cellular uptake by free peptide only[129].

Complexation was measured by Dynamic Light Scattering, abbreviated as DLS. DLS is a method to measure the size distribution of small particles in solution based on their ability to scatter light with a wavelength larger than the particles. The light scattered by these particles will fluctuate with time due to the Brownian motion of these particles in solution. The time scale of these fluctuations can be correlated to the particle size. In principle, not the particle size but the hydrodynamic radius is measured. DLS measurement is also possible for proteins whose size is within the range of DLS. Note that the intensity of a particle peak cannot be directly correlated to the number of particles corresponding to that size.

DLS was measured in a solution of 1 µM peptide concentration in filtered DMEM-CM (Dulbecco's Modification of Eagles Medium-Culture Medium). Compound **4.3** was measured in an overall concentration of 2 µM due to an error in concentration determination. The DMEM contains 10 % (v/v) of FBS. Figure 4.3 shows the size distribution of a mixture of the peptides and cell culture medium separated by a small shift. No complexation for the GCN4 peptide **4.1** occurred because the size distribution in function of volume fraction was similar to that of filtered DMEM-CM. The compounds with a deoxycholic acid scaffold (**4.3** and **4.4**) formed complexes with serum proteins to some extent probably due to their higher hydrophobicity. However, by comparing the volume fraction of these peptides to the one of pure DMEM, it was clear that almost no complexation occurred due to the limited shift of the size distribution (figure 4.3).

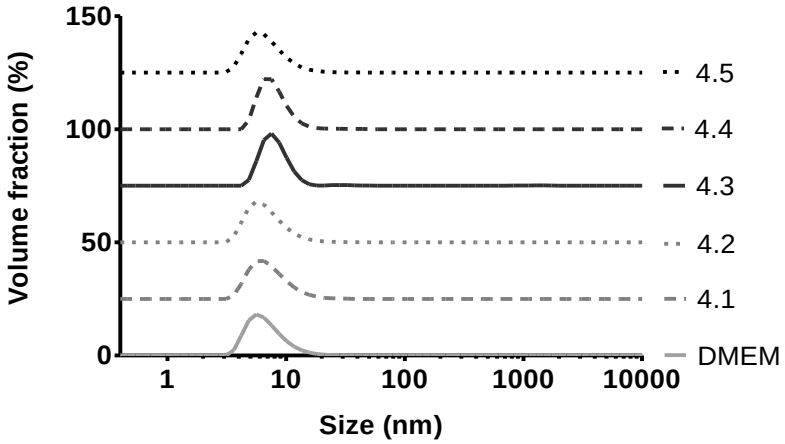


Figure 4.3: Distribution of particle size in solution in function of volume fraction.

In order to further estimate the influence of particles taken up by the cell, fluorescent microscopy was used. A drop of compound **4.3** solution on a cover glass is evaluated by fluorescent microscopy. The presence of a large amount of BSA peptide complex should show a dotted pattern in fluorescence microscopy. In this case, a homogeneous solution is present which means that complexation occurs to a very small extent (figure 4.4). Therefore, it is unlikely that complexation will positively influence the cellular uptake.

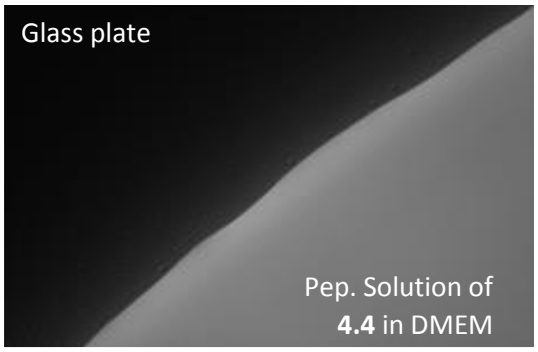


Figure 4.4: Fluorescence microscopy image showing the absence of BSA-protein complexes[1].

4.3 Toxicity: MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) assay

The MTT assay is a colorimetric assay that measures the number of viable cells after incubating the cells with a potential cytotoxic compound[134]. All peptides which were subjected to cell penetration tests were checked for their toxicity toward cells using different concentrations.

The MTT assay is based on the enzymatic reduction of 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide to formazan in living cells (figure 4.5). Several dehydrogenase enzymes are responsible for the conversion of the MTT dye in active mitochondria[134]. Formazan crystals, which are insoluble in water, are generated in proportion to the amount of active mitochondria. In principle, the MTT assay does not measure cell viability as such but metabolic activity within cells represented by the activity of mitochondria. The cell viability is often proportional to the metabolic activity.

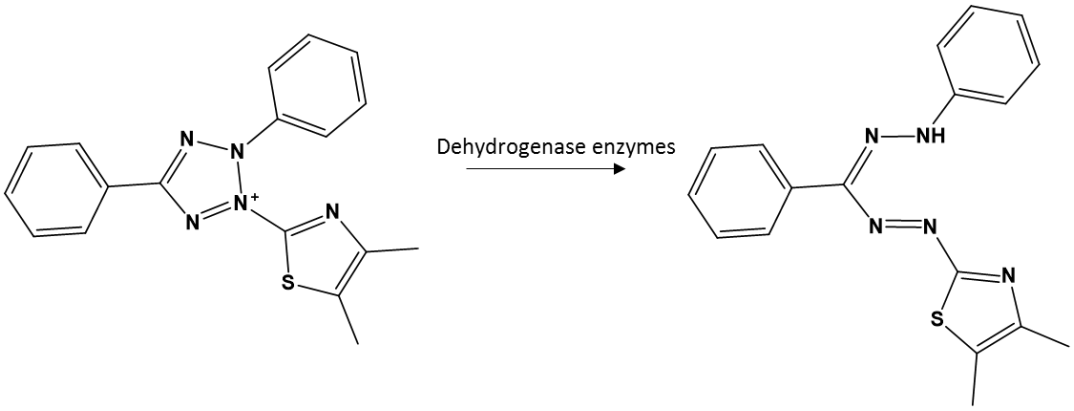


Figure 4.5: Conversion of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium ion to formazan.

RAW264.7 cells cultivated in DMEM (10^4/well, 200 µL) were plated in a 96-well plates and incubated overnight at 37 °C and 5 % CO_2. The peptides and control compounds were added in an overall concentration of 0.25 µM and 1 µM and incubated for 24h under the same conditions. The MTT solution was then added to the aspirated wells and incubated for 3 h. After removal of the cell medium, purple

[1] Reproduced with permission from the master thesis of Dorien Van Lysebetten: "Design and Synthesis of Stapled Peptides for Enhanced DNA Binding and Cell Penetration"[265]

formazan crystals were dissolved in DMSO. Then, UV-measurement at 570 nm was performed with a plate reader to check cell viability quantitatively. An observation with the naked eye already gave a good idea about the toxicity of the compounds due to the disappearance of color in the well. Comparison of the absorbance of the formazan solution of the sample to the absorbance of a positive (incubation with ultrapure water) and a negative control (incubation with DMSO) gave quantitative results of cell viability as in the equation 1:

$$Cell\ viability\ (\%) = \frac{A - A_{pos}}{A_{neg} - A_{pos}} * 100$$

Equation 1: A, A_{pos} and A_{neg} refers to absorbance, absorbance due to water and absorbance due to DMSO respectively.

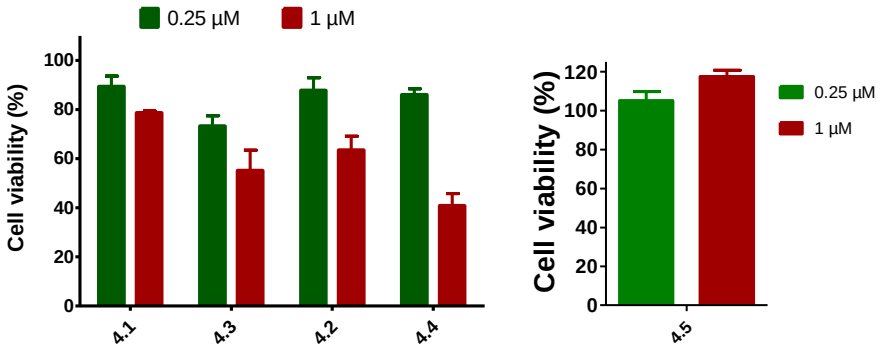


Figure 4.6: MTT assay for all the peptides measured at 0.25 µM and 1 µM.

Cell viability was evaluated according to the ISO10993-5 norms which state that compounds are cytotoxic if the assay points to a cell viability lower than 70 %. Figure 4.5 shows the percentage of viable cells for all peptides. At 0.25 µM, the cell viability of **4.3** was close to 70 %. On the other hand, at 1 µM, the deoxycholic scaffold containing compounds (**4.3** and **4.4**) and the dimerized GCN4 (**4.2**) were cytotoxic. This was not unexpected because bile acids have detergent like properties that can damage the membrane resulting in cell death. A cell viability higher than 100 % which was observed for **4.5** is possible due to higher metabolic activity in the cells. Because the formation of formazan is dependent of the activity of mitochondria, stressing of cells, with foreign compounds, can increase the activity of the cell.

4.4 Cell penetration and localization

4.4.1 Confocal microscopy

4.4.1.1 Description of the technique

Confocal microscopy is an ideal method to investigate the localization of fluorescent peptides in the cell. The laser of the microscope excites the fluorescently labelled CPP which then emits light. A special detector which suppresses out-of-focus light, creates an image of this emitted light. Staining of specific cell

compartments by fluorophores emitting at different wavelengths than fluorescein, allows subsequent detection and localization of these compartments. By merging different emission patterns at different wavelengths, localization of the fluorescently labelled peptide is possible. In current thesis, cells are stained using Hoechst and Cholera toxin subunit B (CTB) Alexa Fluor 647 conjugate to visualize the cell nucleus and cell membrane respectively. The Hoechst dye has specific affinity for the minor groove of DNA[135] and therefore colours the nucleus blue[136]. CTB, on the other hand, has affinity for the cell membrane which colours red due to the conjugated Alexa Fluor 647 dye[137]. The peptides used were labelled with fluorescein[138] which emits green light. Due to the difference in emission wavelength, the three dyes can be detected separately. Processing of the data gave overlay images whereby nucleus, cell membrane and fluorescein connected molecules are shown in blue, red and green respectively.

4.4.1.2 Localization of peptides measured by confocal microscopy

RAW264.7 cells cultivated in DMEM (10^5/well, 300 µL) were plated in confocal plates and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and incubated for 2 h at the same conditions or on ice. Cells were washed and fixated with 2 % of Paraformaldehyde for 30 min at 37°C. Cells were stained by 30 min incubation with a 0.2 % solution of CTB-AF647 and 0.2 % of Hoechst in PBS with 1 % BSA. This was done to visualize the cell nucleus (blue) and cell membrane (red) respectively. Cells were resuspended in PBS and measured with confocal microscope (Leica SP5 equipped with a 63x (1.4 NA) oil immersion objective) at 3 different wavelengths (405 nm, 488 nm and 643 nm). The peptides used were labeled with fluorescein which emits green light. To understand the uptake mechanism better, three samples are made in duplicates and incubated at 4 °C to inhibit active transport. Due to the difference in emission wavelength, the three dyes can be detected separately. Processing the data with the software Image J gave overlay images whereby the nucleus, cell membrane and fluorescein are shown in blue, red and green respectively.

The GCN4 sequences coupled to deoxycholic acid (**4.3** and **4.4**) or coupled to cyclodextrin (**4.5**) were compared to GCN4 peptide control (**4.1**) and the dimerized GCN4 sequence (**4.2**) (figure 4.7 and 4.8). All these peptides were taken up at 37 °C although dimer **4.2** and **4.5** aggregated on the outer cell surface which was indicated by oversaturated green dots outside the red cell membrane (figure 4.7 A and B). This aggregation was less present for the deoxycholic acid peptides (**4.3** and **4.4**). Although all the compounds were taken up at 37 °C, there is clearly an enhanced uptake when deoxycholic acid is used as a scaffold as seen from the mean fluorescence values (figure 4.10 left). The mimic using a cyclodextrin scaffold shows considerably lower uptake than the cholic acid but similar to that of the non-binding GCN4 dimer **4.2**. The monopodal deoxycholic acid derivative **4.3**, also non-DNA binding and more hydrophobic, has properties similar to those of its dimeric counterpart **4.4** in terms of uptake.

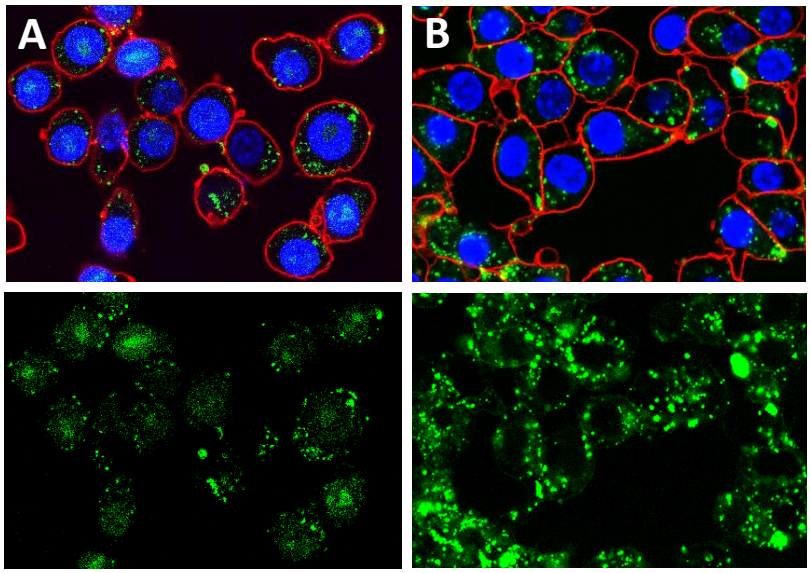


Figure 4.7: Confocal microscopy images of the uptake of (A) **4.1**, (B) **4.3** at 37 °C and 0.25 µM. Upper panels show accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panels show fluorescein image. (Scale bar = 20 µm).

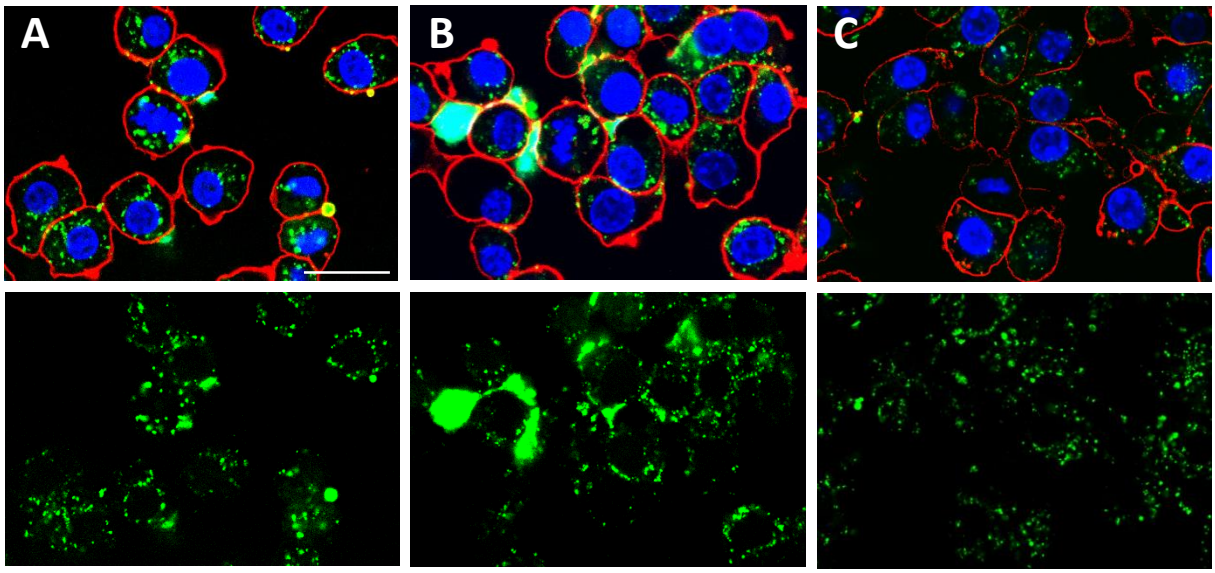


Figure 4.8: Confocal microscopy images of the uptake of (A) **4.2**, (B) **4.5** and (C) **4.4** at 37 °C and 0.25 µM. Upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panel shows fluorescein image. (Scale bar = 20 µm).

4.4.2 Flow cytometry

4.4.2.1 Description of the technique

Flow cytometry is an laser-based analysis technique that consists in measuring light scattering and fluorescence emission of particles or cells in suspension[139]. These particles or cells move one by one in a flow chamber and their scattering properties are measured by a laser beam in combination with measuring the fluorescence light emitted from a dye inside or connected to the cell. Therefore, information about the size and granularity of the cell can be obtained by analyzing the forward and side scattering respectively. In addition, quantitative information can be achieved by measuring the fluorescence emission at a specific wavelength. In general, a flow cytometer consists of a hydraulic system, a set of lasers and detectors and a computer that collects and converts the data from the detector[139]. The hydraulic system ensures the correct arrangement of the cells in order to make them pass one by one through the laser beam. The laser produces monochromatic light that can be used to excite a dye inside or attached to the cell. A histogram plots the emission intensity in function of the counted cells as for example in figure 4.9.

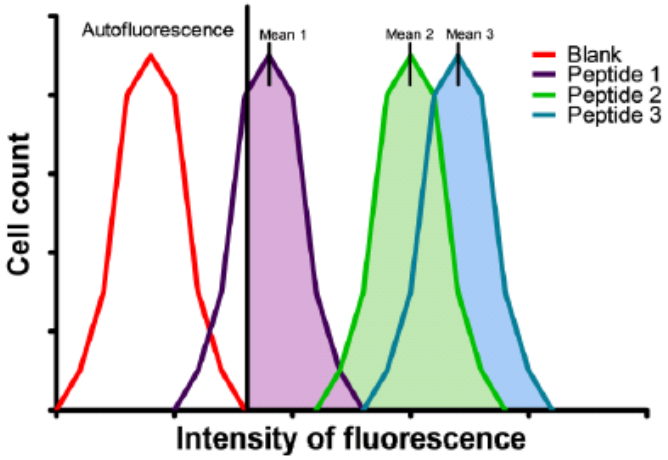


Figure 4.9: Principle of data representation of flow cytometry. A threshold is established to separate the fluorescence emitted by the cells (autofluorescence) from the one emitted by cells containing fluorescently labeled peptides. For example, peptide 1 is not taken up by all cells, as the area under the curve is in the autofluorescence zone and the percentage of peptide positive cells is lower than 100 % proportional to the colored area under the curve. Peptide 3 was taken up better due to its higher mean fluorescence value (Mean 3), although the percentage of peptide positive cells for both peptide 2 and 3 is 100 %. It is therefore important to both compare the mean fluorescence and the amount of peptide positive cells[3].

RAW264.7 cells cultivated in DMEM (10^5/well, 1 ml) were plated in a 24-well plate and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and 1 µM and incubated for 2h at the same conditions or on ice. Cells were washed with PBS and detached with

[3] Reproduced with permission from the master thesis of Dorien Van Lysebetten: "Design and Synthesis of Stapled Peptides for Enhanced DNA Binding and Cell Penetration"[265].

Na_4EDTA. Cells were re-suspended in PBS and added to the BD Accuri flow cytometer. The amount of peptide associated to the cell was measured by flow cytometry. Experiments were carried out in duplicate. In order to control the fluorescence of the background, blank cells were measured and the signal received corresponded to auto-fluorescence. This allowed the establishment of a threshold to separate cells emitting light due to auto-fluorescence from cells containing fluorescently labeled peptides. Thus, the number of cells which emit sufficient intensity of fluorescence due to the presence of fluorescently labeled peptides could be achieved by measuring the area under the histograms beyond this threshold. This number was further referred to as the percent of peptide positive cells. As an example, by measuring two different peptides, they can have 100% of peptide positive cells, but one can be better uptaken in the cell due to its higher mean fluorescence value. If a peptide is not taken up by all cells, the percentage of peptide positive cells is lower than 100 % proportional to the colored area under the curve. It is therefore important to compare both the mean fluorescence and the amount of peptide positive cells.

4.4.2.2 Analysis of the peptides by Flow Cytometry

During flow cytometry measurements no comparison was possible between the results of single fluorescently labeled peptides and the peptides containing two fluoresceins. Therefore the single fluorescently labeled peptides (**4.1** and **4.3**) were discussed separately from the ones containing two fluoresceins (**4.2, 4.5** and **4.4**).

A concentration dependent uptake was observed for all peptides. Compound **4.3** which is a single GCN4-strand grafted on a deoxycholic scaffold had substantially higher uptake (figure 4.10). These findings emphasize the importance of conjugating peptides to a bile acid.

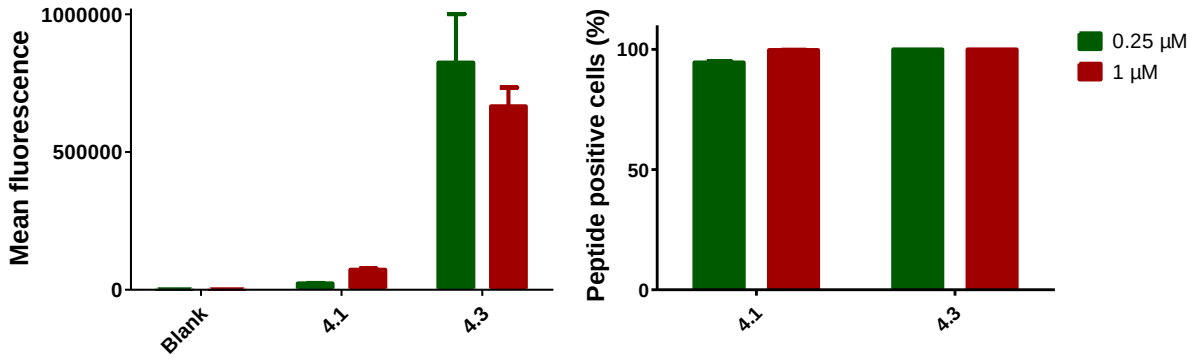


Figure 4.10: Mean fluorescence (left) and peptide positive cells (right) of third test of single fluorescently labeled peptides.

The increased uptake of **4.3** at 0.25 µM and 1 µM compared to the other peptides was not unexpected. However, a decrease in mean fluorescence was observed when switching from 0.25 µM to 1 µM. The tendency of a bile acid to disrupt the cell membrane due to its detergent like properties can be responsible

for a maximum uptake level. However, the reason for the decrease in the mean fluorescence when increasing the concentration from 0.25 µM to 1 µM is unknown.

As mentioned before the uptake of the peptides labeled with two fluoresceins was interpreted separately from the other peptides. In figure 4.11, mean fluorescence and peptide positive cells are shown for these doubly fluorescent labeled peptides. The uptake of peptides **4.4** and **4.5** which contain deoxycholic acid or cyclodextrin scaffold coupled to two peptide strands, was compared to the dimerized GCN4 peptide **4.2** which was used as control. All these peptides labeled with two fluoresceins were taken up by 100 % of the cells at both concentrations. The mean fluorescence value of dipodal cholic **4.4** was considerably higher compared to dimer **4.2** as was also observed for peptide **4.3** compared to the control **4.1**. Also here, the mean fluorescene value did not increase as much as expected upon switching to the higher concentration which can indicate a maximum uptake level of the peptides coupled to deoxycholic acid. Coupling of a cyclodextrin scaffold did not enhance the uptake (figure 4.10, left). This was not an unexpected result and once again established the importance of the bile acid in cellular uptake.

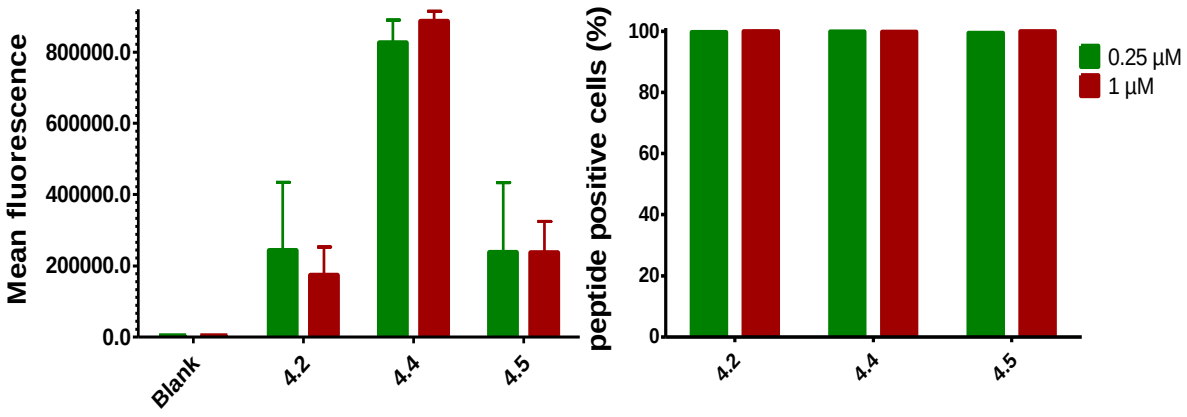


Figure 4.11: Mean fluorescence (left) and peptide positive cells (right) as a result of the first test with doubly fluorescent labeled peptides.

4.5 Description and studies of uptake mechanisms

4.5.1 Endocytosis

The mechanisms through which Cell Penetrating Peptides (CPPs) are imported into the cell have been of interest since more than 25 years ever since the discovery of the trans-activating transcriptional factor (TAT) CPP in 1988[140]. Due to the enormous complexity of these mechanisms, in reality, we are far away from explaining what truly happens to a peptide *in cellulo*. However, a general explanation has been provided in this section. From the previous studies on cell uptake described in literature[141,142], we can conclude that one of the major pathways through which CPPs interact with the cellular membrane is by inducing an endocytic uptake by the enclosure of the CPPs in a vesicle. This vesicle is further pinched off

from the plasma membrane and release into the cytosol, generating an intracellular organelle, called endosome. The release of the CPPs from the endosome to the cytosol is caused by the acidification of the vesicle, allowing the CPPs to interact with the target[141].

Several internalization pathways have been described (figure 4.12) and they have been classified depending on the size of the particle taken up by the cell. These mechanisms are named as phagocytosis and pinocytosis. The first one is related to the uptake of large particles, for the elimination of pathogens and for apoptotic and infected cells[141–143]. On the other hand, pinocytosis is involved in the internalization of small particles such as fluids, solutes and components of the cellular membrane. It comprises a subgroup of four different mechanisms: (1) macropinocytosis, (2) clathrin-mediated uptake, (3) caveolae-mediated uptake and (4) clathrin-/caveolai independent uptake. The last three mechanisms are named as raft-dependent endocytosis, as they are supported by the presence of lipid rafts. These lipid rafts are part of the cell membrane with high composition in cholesterol, sphingolipids and receptors[142].

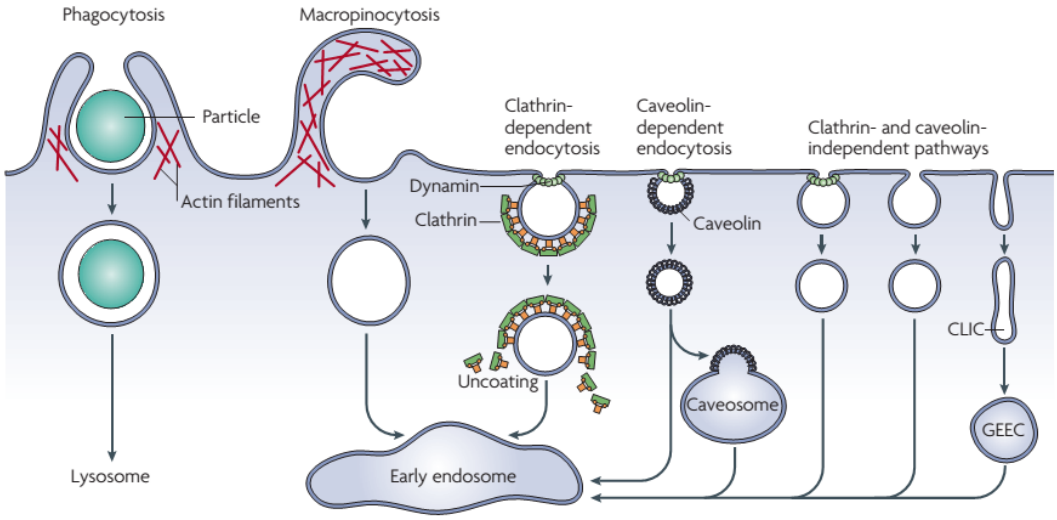


Figure 4.12: Different endocytic uptake mechanisms[142].

During macropinocytosis, the plasma membrane bends giving rise to curved ruffles, that when shielded, forms large vesicles called macropinosomes. Due to this mechanism, the cell is able to internalize big volumes of fluids and dissolved particles, such as viruses, bacterias and cellular fragments[142,143]. One of the modes of uptake of arginine rich peptides is by macropinocytosis as described by Wadie and coworkers. The escape from these type of vesicles is achieved by lowering the pH to a point at which the peptides degrade, thereby leaking out of the internalized components[144]. CPPs that are taken up by macropinocytosis are not further internalized to lysosomes (vesicles specialized in breaking down biomolecules), preventing their degradation[145].

A second type of mechanism is the clathrin-mediated endocytosis (CME), which is the most extensively studied and best understood. The CPPs that go through this mechanism of uptake firstly interact with a receptor located at the extracellular surface, generating a clathrin coated pit[141,146]. When it is pinched off, a

clathrin-coated vesicle (CCV) is formed together with the removal of the clathrin protein. These uncoated vesicles are then fused to form an early endosome. Studies have revealed the importance of CME in the regulation of signal transduction (i.e. the activation of a specific receptor located inside the cell or on the cell surface by an extracellular molecule[147,148]), as it is related to the number of receptors at the outer surface of the membrane.

In case of Caveolae-mediated endocytosis, like in the previous two mechanisms, flask-shaped invaginations, called caveolaes are formed[141,142,149]. Caveolaes contain caveolin proteins and lipid rafts like cholesterol and sphingolipids. Although this mechanism is still not well understood, it has been shown that it participates in cell adhesion and anchoring-dependent cell growth[141,142].

The last class, clathrin-/caveolae- independent endocytosis, does not fit within the description of the previous uptake mechanisms[150,151]. The presence of uncoated invaginations on the surface of the cell membrane encouraged the investigation of a possible alternative endocytic mechanism. Although it has been classified separately, there is still a lack of information and understanding required to fully understand both the molecular basis of membrane deformation and scission and the different pathways and their individual functions in their physiological contexts.

Studies performed in live cells demonstrated that CPPs are taken up by endocytotic pathways such as macropinocytosis, CME and caveolin-mediated endocytosis although other mechanisms cannot be completely discarded[141]. The specific uptake mechanism of CPPs is determined by the use of specific markers, uptake inhibitors and knock-down cells. However, due to the dependence of the mechanism on the type of cell line used, the obtained results are not entirely specific for one mechanism. For example, the inhibition of one pathway can enhance the effect of others and therefore can result in one drawing the wrong conclusions. In addition, other factors such as incubation time, concentration and physiochemical parameters restrict the access to general conclusions about the uptake of CPPs[143].

The most common CPPs for the transport of cargo through the cell membrane are the TAT-peptide, polyarginine and penetratin. Their uptake mechanisms have been evaluated and are understood better despite the differences based on the variations in cargo[152–154]. In general, it is documented that the Tat-peptide is taken up by macropinocytosis, clathrin-mediated, independent of clathrin and even caveolae-mediated endocytosis[128]. On the other hand, the uptake of polyarginine occurs via micropinocytosis and the length of the sequence is crucial for internalization[141]. In both cases, binding to the heparin sulfate proteoglycans on the outer face is essential for cell uptake. However, in the case of polyarginine, there are indications that other mechanisms are involved during internalization of the peptide, such as the inability of polyarginine R16 to penetrate[128,141]. Finally, penetratin mainly penetrates via clathrin-mediated endocytosis or clathrin-independent and lipid raft-dependent endocytosis.

Due to all the parameters described above which affect cell uptake, namely the type of peptide and cell line and concentration of the peptide it is not possible to draw conclusions about the exact mechanisms of CPPs internalized in the cell. However, the ongoing research on specific inhibitors and markers for each of the

pathways should elucidate and unravel the factors involved in each mechanism, as well as the dependence of the cargo on the cell uptake process.

4.5.2 Transduction

Despite the importance of the endocytic pathway, there are peptides internalized by mechanisms that do not require the formation of endosomes. These mechanisms are grouped under the name transduction and occur when the physical-chemical properties of the peptides stimulate interaction with the cell membrane[155]. Transduction is also defined as direct translocation, as it is independent of energy consumption and therefore it is not inhibited upon incubation at 4°C. This is contrary to endocytosis which needs a higher temperature for the active uptake. Recently, three different pathways for transduction have been described by the group of Copolovici: the inverted micelle model, the carpet model and pore formation[143].

The inverted micelle pathway has been proposed for the penetratin family[156]. This mechanism consists of the interactions between the positively charged residues of the peptide with the negatively charged heads of the phospholipids. The hydrophobic residues are responsible of the formation of the inverted micelle by membrane destabilization. The polar groups of the peptide helix point inside the micelle. The peptide is then transported through the non-polar core of the bilayer in an inverted micelle, before reaching the polar intracellular environment (figure 4.13 A). Due to the lack of hydrophobic residues in the sequence of tat and polyarginine peptides, this pathway cannot explain their uptake mechanism. The carpet model is based on the initial interactions between the peptide and the lipid head groups that lead to rotation of the helix, permitting the non-polar residues to establish contact with the lipid tails. As a consequence, the membrane bilayer is disrupted and the peptide can penetrate in the cell. This process is favored by the presence of guanidinium groups, such as in the case of polyarginine and tat peptides (figure 4.13 B). The last mechanism, pore formation, is generated by the aggregation of several α-helices containing an amphipathic phase. The hydrophobic part of the helices interact with the membrane forming a tunnel, in which the hydrophilic parts are oriented into the tunnel. In this model, the peptides stay inserted in a transmembrane orientation and in an associative state to form a cylindrical structure or ‘pore’. There are two types of pores that can be formed by this pathway, the barrel-stave pore (figure 4.13 C left) or the toroidal pore (figure 4.13 C right).

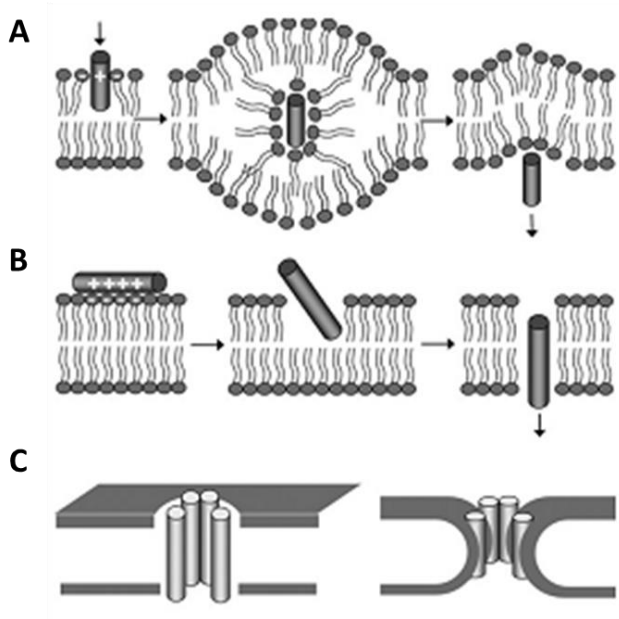


Figure 4.13: Model of transduction of peptides[157].

Apart from the three transduction mechanisms explained above, the presence of other pathways has also been observed. Although these are not well characterized, definitions have been reported in literature. Electroporation, a well-established physical method of disrupting cell membranes and consequently allowing the permeabilization of CPPs by their interactions with the membrane or by internalization via weak spots in the membrane is one such example. Another described method by the group of Bechara[128] is adaptive translocation in which positively charged and arginine-containing oligomers can form membrane-soluble ion pair complexes by association with negatively charged, bidentate hydrogen bond acceptor groups of endogenous membrane constituents. A decrease of the polarity of the ion-pair complexes when inserted in the cell membrane causes the partition of the lipid bilayer and the migration of the peptide towards the cytosol. The rate of migration depends on the membrane potential. Finally, the complex dissociates in the inner part of the membrane and the peptide enters the cytosol. Although these examples have been documented in literature, the ambiguous proof-of-concept for these mechanisms and the dependence on the peptide used and the concentration for these mechanisms causes uncertainty about their validation.

In general, it should be noted that independent of the uptake mechanism, the first interaction arises when the CPPs interact electrostatically with the negatively charged proteoglycans, phospholipids and less commonly proteins. Although the importance of these negatively charged groups on the outer surface is not denied, their further influence on the uptake is still unclear. Another important aspect independent of the uptake mode is the concentration of CPPs needed to give a biological response. Ideally, this should not exceed 10 µM to ensure one has a therapeutically useful concentration for further testing[158].

4.5.3 Endosomal escape

The endocytic processes explained in the section 4.5.1 require the formation of endosomes where the CPPs can be internalized and then taken up into the cell. However, in certain cases CPPs need to be released from the endosomes to reach their final localization (ex. the cytoplasm or the nucleus). Despite the importance of endosomal escape for the uptake of CPPs, very little is understood about the parameters that control it[128]. In order to observe endosomal leakage, polyarginine R6 and Tat peptides have been linked to a fluorescein isothiocyanate (FITC), allowing the access to privileged information about the release of peptides from endosomes[159].

Once the CPPs are internalized in the endosomes, they should be released to be able to exert their function inside the cytoplasm or nucleus. The endocytic pathway is based on the formation of early to late endosomes with an internal pH of 5 that will fuse with lysosomes containing digestive enzymes afterwards[160] (figure 4.14). This can result in degradation or capture of the peptide before it has had the chance to reach its desired target and perform its function. Therefore, endosomal capture is considered as a problem in cellular delivery. Thus, several strategies have been developed to avoid the sequestration of CPPs inside endosomes and enhance their bioactivity. These are based on mechanisms that allow endosomal escape, such as pore formation in the endosomal membrane, conformation changes on the enhancers of endosomal escape and pH buffering effect[161]. The endosomal-releasing agents that facilitate these methods are classified into viral and bacterial proteins and synthetic biomimetics for nucleic acid and protein delivery among others. However, the usage of these systems is limited due to stability problems. It is established that an optimal delivery agent should be efficient and non-toxic.

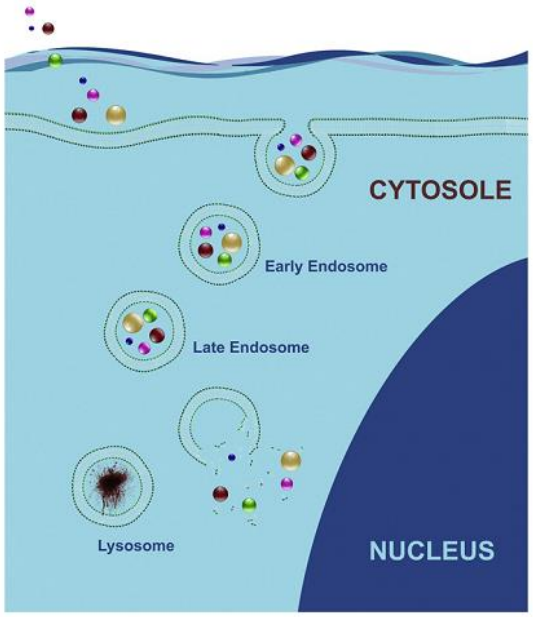


Figure 4.14. Representation of the internalization of therapeutics into the cell through endocytosis and subsequent endosomal escape[162].

As an example of this strategy, the group of Bernard Lebleu and cowrokers have modified Penetratin via an elongation of the N-terminus by varying numbers of arginines for the delivery of PNA to the nucleus in the absence of endosomolytic agents[163]. From the various peptides synthesized, the R6Pen peptide (RRRRRRRQIKIWFQNRRMKWKKGGC) is the most active towards nuclear uptake. Other peptides undergo endosomal escape by changing their conformation depending in the pH, the subsequent endosomal membrane disruption and the final leakage of the contents. These peptides are called fusogenic and they are part of the N-terminal sequence of the influenza virus hemagglutinin subunit HA-2[164].

4.5.4 Uptake mechanism tested by flow cytometry

Due to the limited knowledge and understanding about some uptake mechanisms, research into distinguishing the different mechanisms is highly complicated. In addition, the overall interplay and interconnection between the mechanisms and the absence of highly specific inhibitors or markers which on itself will not influence the experiment, impede the determination of a final conclusion. On the other hand, the well-known distinction between the two main mechanisms, endocytosis and transduction, can be investigated by simultaneous incubation at 4 °C (or lower) and 37 °C. Therefore, half of the 24-well plates, which were plated the day before, were incubated on ice to make sure the temperature of cell culture medium, cells and well plate was below 4 °C. The other half was incubated at normal conditions. Then the peptides were added and the plates were incubated at 4 °C and 37 °C for 2 h, on ice or in the oven respectively. The cells were then washed, detached from the plates and measured by flow cytometry. In order to process the results and to create histograms, plot overlays and quantitation, we used FlowJo, a software commonly used to analyze flow cytometry data and that can be employed independently of the instrument used. GraphPad Prism, a commercial scientific 2D graphing and statistics software was also used for statistical analysis.

The following parameters were used while performing cell cultivation: DMEM (105/well, 1 ml) plated in 24-well plates and incubated overnight at 37 °C and 5 % CO2. The peptides were added such that the overall concentration was 0.25 µM and 1 µM. They were then incubated for 2h in the same conditions mentioned above or on ice. The cells were washed with PBS and Na4EDTA was used for the detachment of the cells. They were then re-suspended in PBS and added to the BD Accuri flow cytometer. All experiments were repeated twice to check for consistency. To measure the amount of peptide associated to the cells flow cytometry was used. Blank cells were measured and the signal received through those cells corresponded to auto-fluorescence. A suitable threshold was chosen to distinguish the auto-fluorescent cells from those fluorescently labelled. Care was taken to ensure that the area under the histograms beyond the threshold chosen contained sufficient cells and provided enough intensity of fluorescence to label the peptides present in those cells. This number is referred to as the percent peptide positive cells. Two factors have to be considered in case of cell uptake: peptide positive cells and the mean fluorescence. It is quite possible that two peptides may have 100% peptide positive cells but one of them could be better if it had a higher mean fluorescence value. If the peptide positive cells is lower than 100% the percentage of peptide positive cells is proportional to the area under the curve.

Once again a concentration dependent uptake was observed for both monomeric peptides **4.1** and **4.3** and compound **4.3** had substantially higher uptake than monomeric GCN4 peptide **4.1** irrespective of the temperature or concentration (figure 4.13). The enhancement on cell uptake at 37°C for **4.3** was significantly higher than for **4.1**, revealing the influence of a bile acid in active uptake.

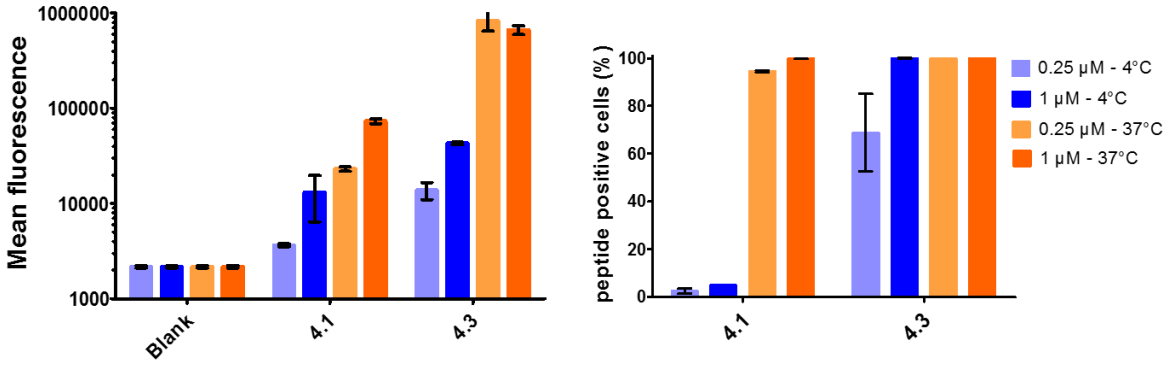


Figure 4.15: Mean fluorescence (left) and peptide positive cells (right) as a result of incubation at 4 °C and 37 °C at a concentration of 0.25 µM or 1 µM.

For the dimeric constructs, the first test was performed at 0.25 µM only. Once again, peptides labeled with one fluorescein (**4.1**, **4.3**) could not be compared with peptides labeled with two fluoresceins (**4.2, 4.4, 4.5**). The quantitation of the uptake was visualized using the mean fluorescence and the percent of peptide positive cells at the two temperatures is as depicted in figure 4.15 for the singly fluorescent labeled peptides.

The dipodal GCN4 mimic **4.4** was internalized in all cells at both 4 °C and 37 °C which is depicted in figure 4.16. Peptide **4.2** was also taken up both passively and actively whereby at 37 °C both the pathways were possible which resulted in an increased mean fluorescence at 37 °C compared to the mean fluorescence at 4 °C. At 4 °C, the dimerized GCN4 peptide **4.2** was only present in 20 % of the cells where at 37 °C all cells were 'dimer **4.2**-positive'. This peptide was probably translocated via endocytic pathways which also resulted in a lower mean fluorescence value at 4 °C. An important aspect is the lower cell membrane flexibility at low temperature which can result in decreased passive uptake at 4 °C compared to passive uptake at 37 °C[128]. In this study, this was not taken into account due to the difficulty in measuring the influence of cell membrane flexibility.

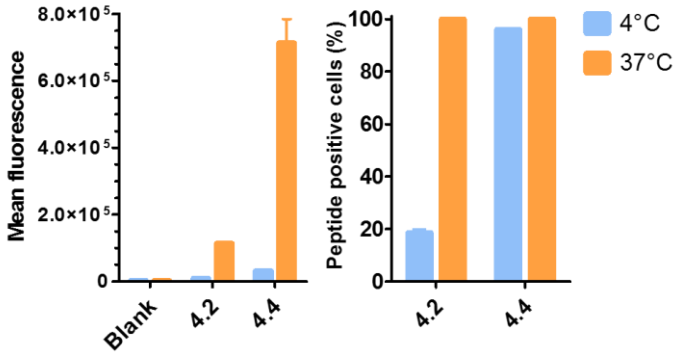


Figure 4.16: Mean fluorescence (left) and peptide positive cells as a result of incubation at 4 °C and 37 °C of doubly fluorescent labeled peptides at a concentration of 0.25 µM.

The dependence of the number of arginines on the transduction mechanism was described by Lättig-Tünneman et al. In addition, they also studied the concentration-dependent threshold which exists for this mechanism[165]. A low micromolar range threshold usually points to an endocytotic pathway, however, this may not always be the case and may not always be known either[158]. For our specific case, we decided to check for a trend based on the theory proposed by Lättig-Tünneman by repeating the tests using concentrations of 0.25 µM and 1 µM. Due to the limited material at hand further increasing the concentration was not feasible. However, according to Tünneman, it is quite possible that the threshold concentration was exceeded when the test was performed at 1 µM as the threshold range is between 1 and 10 µM[166].

The monopodal GCN4 mimic **4.3** and the dipodal GCN4 mimic **4.4** proved to benefit from an increased uptake compared to the GCN4 peptide **4.1** and the GCN4 dimer **4.2** respectively (figures 4.15 and 4.16). Even at 4 °C and a 1 µM concentration, almost a 100% of cells contained peptide **4.3**. The same was true at 4 °C and 0.25 µM, although this time the uptake was not 100 % but still significant. Therefore, monopodal mimic **4.3** may be taken up both passively and actively which was indicated by the increased mean fluorescence value at 37 °C compared to 4 °C, similar to the observations for peptide **4.4** in the first test (figure 4.16). To confirm the uptake at 4 °C, confocal microscopy is needed.

4.5.5 Uptake mechanism tested by confocal microscopy

The control peptide was not internalized at 4 °C as can be seen from figure 4.17 A. This points to the uptake via an active transport mechanism. Remarkably for compound **4.4** at 4 °C only insertion in the cell membrane was observed which is displayed in figure 4.17 A. The yellow dots on the cell membrane are a combination of red signal from the cell membrane and green signal from fluorescein (figure 4.17 A, upper panel). Due to accumulation in or onto cell membrane, it was possible to measure 100 % peptide positive cells and a high mean fluorescence in flow cytometry observed in figure 4.16. However, at 37°C, whereby both passive and active uptake is possible, the uptake is considerably higher (figure 4.16). This proved that the deoxycholic acid coupled peptides were also internalized via active transport. Passive transport was not

possible because the peptides stayed in the cell membrane or on the outer cell surface. Unfortunately, no peptides were localized inside the nucleus. This absence could be a result of rapid degradation of the peptides, endosomal entrapment or the absence of a NLS.

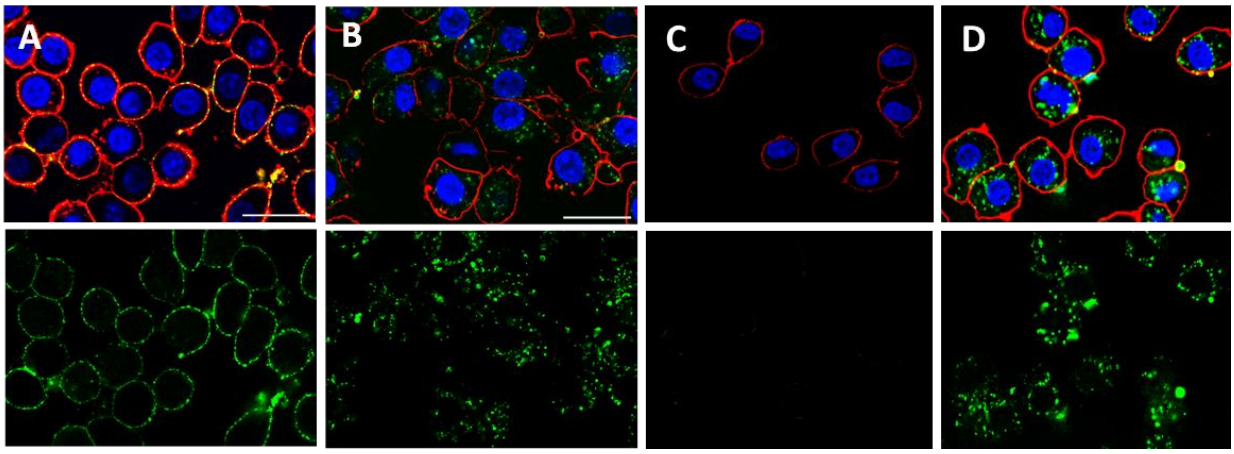


Figure 4.17: Confocal microscopy images of the uptake of **4.4** at 4 °C (A) and at 37 °C (B) and **4.2** at 4 °C (C) and at 37 °C (D). Upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panel shows fluorescein image in case of A and B. (Scale bar = 20 μm).

The confocal microscopy confirmed that the GCN4 dimer **4.2** has almost no uptake at 4 °C (figure 4.17 C) but is taken up by the cells at 37°C (figure 4.17 D) indicating an endocytotic pathway.

4.6 Conclusion & Perspectives

The synthesized scaffold-based mimics of the GCN4 leucine zipper TF that showed sequence specific DNA binding towards the CRE sequence were tested for DNA binding in vivo. For the compounds that have affinity for the DNA, only those peptides bearing a steroid scaffold were able to penetrate the cell membrane. The CD conjugate showed accumulation in and around the cell membrane rather than uptake in the endosomes. The moiety responsible for the high mean fluorescence even at 4°C is the steroid nucleus. However, studies on the uptake mechanism of **4.4** revealed that the internalization occurred via active transport despite the high amount of peptide positive cells.

Although all the compounds were taken up at 37 °C (figure 4.16 and 4.17), there is clearly an enhanced uptake when deoxycholic acid is used as a scaffold as seen from the mean fluorescence values (figure 4.15 left). The mimic using a cyclodextrin scaffold **4.5** shows considerably less uptake than **4.4** but similar to that of the non-binding GCN4 dimer **4.2**. The monopodal deoxycholic acid derivative **4.3**, also non-DNA binding and more hydrophobic has properties similar to those of its dimeric counterpart **4.4** in terms of uptake (figures 4.7 and 4.8). The most interesting results were obtained when comparing the localization of the DNA binding cholic acid dimer **4.4** at 4°C and 37°C (figure 4.17). At 4°C, where active transport is inhibited, only insertion in the cell membrane was observed. Accumulation in and around the cell membrane resulted in a high mean fluorescence and high percentage of peptide positive cells (figure 4.16 left). However, at 37°C, where both passive and active uptake are possible, the uptake is considerably higher.

This indicates that the deoxycholic acid coupled peptides **4.3** and **4.4** are internalized via active transport. Despite the enhanced uptake shown by **4.4** there is no uptake in the nucleus. We already know that increased uptake of peptides bearing arginine residues is observed due to the bidentate character of the guanidinium groups. The cell surface contains negatively charged groups like phosphates and carboxylate groups. From the behaviour of compound **4.4** (figure 4.17A) one can conclude that deoxycholic acid prefers membrane association. Since this interaction is also prevalent at 4°C, unlike the other peptides, **4.4** shows uptake at 4°C. Compound **4.4** is accumulated in and around the cell membrane which is most likely due to passive transport.

Despite these results no peptides were localized inside the nucleus. We believe that coupling a short NLS to the peptides or the scaffold can perhaps increase nuclear localization granting access to in cellulo DNA binding. The most suitable attachment point for this would be the modification of the carboxylic acid at C24 in the steroid scaffold. Only a few of these bioactive CPPs are known today[143]. It is therefore important to fully exploit the possibilities of these bioactive CPPs by further research into stability, cell penetration and localization of these peptides.

CHAPTER 5

FROM ARTIFICIAL HOMODIMERIC TF MIMICS TO A HETERODIMERIC TF MIMIC

Synthesis of GCN4 bZIP transcription factor models has been successfully achieved with the steroids and cyclodextrins as keystones for substitution of the dimerization domain of the TF. Using the gained expertise on simplifying proteins by chemical tools, we decided to attempt the miniaturization of the cMyc/Max bHLH transcription factor, whose role in the development of tumors make it an interesting target towards gene therapy. By broadening the scope of our synthetic toolbox we can target different DNA sequences and mimic the various complex, natural TFs.

5.1 Introduction

5.1.1 Description of c-Myc/Max bHLH Leucine Zipper transcription factor

Over the last few decades, development in areas such as biotechnology and chemical biology have led towards new directions in anti-cancer therapy research. Among the promising approaches that have gained attention recently, the inhibition of oncogenes and/or the reactivation of tumor suppressor genes are of particular interest. This is mainly due to the fact that they are responsible for encoding transcription factors that control gene expression and signaling pathways in cancer. They are also related to the maintenance and progress of tumor growth[167]. During cancer development, these transcription factors suffer from an aberrant activation which has consequences on the tumor formation and metastasis. On the other hand, as they are inactive under normal physiological conditions, they can be regarded as desirable targets in gene therapeutic strategies.

One of the major families of transcription factors related to human cancer and consequently targeted for cancer therapy is the Myc/Max/Mad network[168]. These oncoproteins are heterodimers composed of proteins from the Myc, Max and Mad families. Myc-Max binds DNA in a sequence-specific manner and its function is the transcription from promoters containing the sequence CACGTG. The disregulation of the c-Myc gene causes its overexpression which is related to the progression of diseases such as Burkitt's lymphoma, cancer in infancy in the form of neuroblastomas and lung, breast and colon carcinomas. The oncogenic biological activity of the c-Myc protein is regulated by its ability to establish protein-protein interactions with its partner Max to form a heterodimeric transcription factor. In contrast, Max forms stable homodimers that bind the same DNA sequence but do not activate transcription, thus antagonizing the activity of the cMyc-Max heterodimer by competing for the same target sequence. The heterodimer formation is controlled by

the interactions at the leucine zipper domain. While the Max homodimer is formed when a Gln91-Asn92-Gln91-Asn92 tetrad occurs near the C-terminal end of the zipper region, the heterodimer creation is regulated by the Arg423-Arg424-Gln91-Asn92 tetrad establishment (figure 5.1).

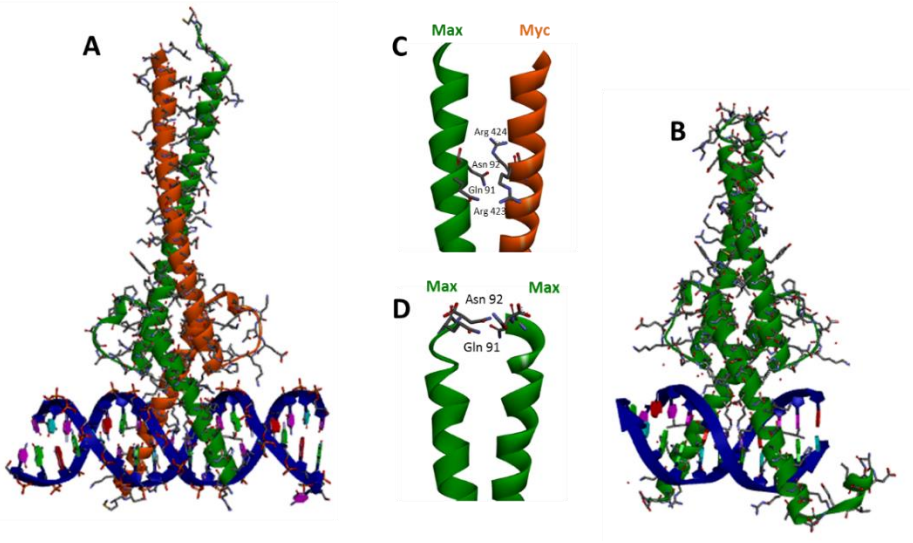


Figure 5.1: Structures of the (A) Myc-Max Heterodimer (B) Max-Max Homodimer and interactions between the C-terminal helix region in case of (C) Myc-Max and (D) Max-Max.

5.1.2 Towards synthetic mimics

Due to the relevant role of this oncoprotein in cancer development, Kent et al[169] made a first approach in the form of the total synthesis of the cMyc-Max protein. Unlike the natural protein Kent synthesized a covalently-linked dimer mimic by convergent chemical ligation of unprotected peptide segments. This approach was chosen as the obtention of a heterodimer composed by Myc and Max counterparts was hampered by the favoured Max/Max homodimeric formation at the supraphysiologic protein concentration. Therefore, Kent proposed the total chemical synthesis of the protein in this manner as an alternative to the construction of native Myc/Max proteins (figure 5.2). In addition, this approach permitted the introduction of structural modifications such as introduction of unnatural amino acids and point mutants and allowed the replacement of E. coli expression systems that are frequently related with protein instability and toxicity.

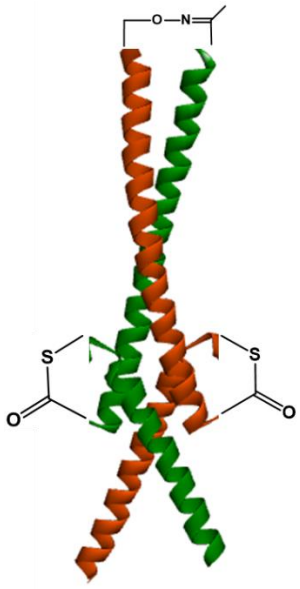


Figure 5.2: Structure of Myc/Max bHLH protein mimic synthesized by Kent. The green ribbon represents the residues that constitute the Max peptide, while the orange ones refers to the Myc peptide. Oxime ligation and thioester formation allow connecting the two different fragments.

The biological functions of the cMyc-Max bHLH ZIP transcription factor as an oncoprotein have not been deeply investigated. One of the reasons could be the lack of good synthetic routes or ligation strategies which could provide sufficient material to investigate the functions of the protein in detail. Unfortunately, the simplification of this construct has not been deeply investigated either. On the other hand, despite sharing the same mode for DNA recognition, artificial models of bZIP transcription factors have been developed by several groups[56–58,127,170] as presented in chapter 2. The eukaryotic bZIP and bHLH-ZIP motifs have one of the simplest protein structures capable of sequence-specific DNA recognition on the major groove. The structural motif of bHLH is very similar to the bZIP. They are both characterized by a dimerization domain, the basic Helix-Loop-Helix-Zipper, which is a prerequisite for DNA binding. They both also havea DNA recognition domain – the basic region Max (Ala701-Asp716) and Myc (Asn500-Asn515), which contain the α-helical peptides responsible of the specific DNA binding. The main difference is, however, the interface between the two domains. In case of the bHLH proteins, the α-helices are interrupted by a loop, while in bZIP they are continuous (figure 5.3).

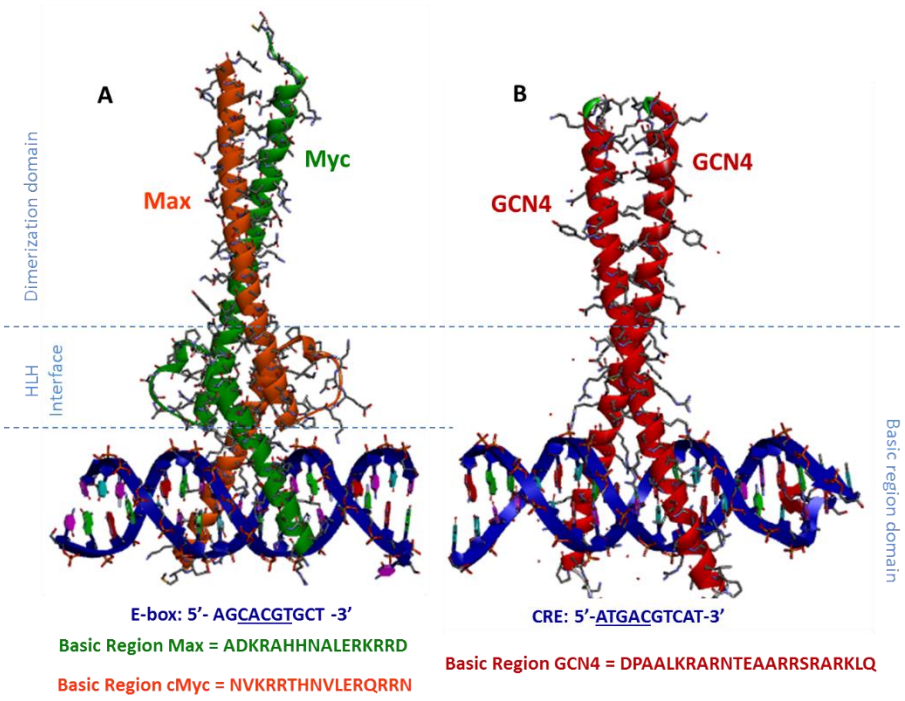


Figure 5.3: Crystal structures of the (A) cMyc-Max protein and (B) GCN4 protein bound to DNA.

Despite the lack of chemical models for the cMyc/Max TF, a successful mimic of the MyoD-MyoD bHLH-type TF was reported by the Morii group[171]. This encouraged us to modify our successful peptidosteroid conjugation strategy to synthesize a heterodimeric counterpart. In our group we have previously attempted to synthesize artificial mimics of this oncoprotein, however, with unsuccessful DNA binding[172]. As a conclusion in the article, it was proposed that a spacer between the steroid-like dimerizer moiety and the basic region of the peptides would facilitate the appendage of the peptides in the correct position to bind to the major groove. Therefore, we report here the synthesis of an optimized model of bHLH ZIP transcription factor that binds specifically to its nonpalindromic target DNA sequence.

5.2 First design developed in the OBCR group

Our first model was developed during the PhD thesis of Dieter Verzele[78] and was based on the decoration of an orthogonally protected diamine deoxycholic acid-based scaffold, C3α-NHAlloc, C12α-NHBoc-diamino-5β-cholan-24-oic acid[48,74] with the cMyc and Max basic region peptides at the C3 and C12 positions respectively synthesized by linear Fmoc/tBu SPPS[79]. As a result of orthogonal N protection, a unique feature of the applied scaffold is the straightforward solid-phase generation of homo- as well as heterodimeric peptidosteroid tweezers. As reported in chapter 2, steroidal building blocks have been broadly studied in drug design due to their amphiphilic nature, bioavailability and trend to enhance peptide biostability. More importantly the properties they possess are vital and related to the design of transcription factor models, i.e. they have a rigid core structure, the correct dimensions for the appendage of peptides which allow major groove binding and they also increase helicity of appended peptides. For synthetic ease, the natural spacer and carboxylic acid functionality at C24 allowed the anchoring of the scaffold on

TentaGel solid support and the development of a strategy on solid phase. UV irradiation of the Holmes' photolabile linker (nitroveratryl-based photosensitive linker 2 (4-{4-[1-(9-fluorenylmethoxycarbonylamino)ethyl]-2-methoxy-5-nitrophenoxy}-butanoic acid)[88] was chosen as a cleavage strategy due to its compatibility with other functional groups and orthogonality to the conditions used during peptide synthesis and to Alloc and Boc deprotections. The desired compound could thus be released as a primary carboxamide at C24 (Figure 5.4).

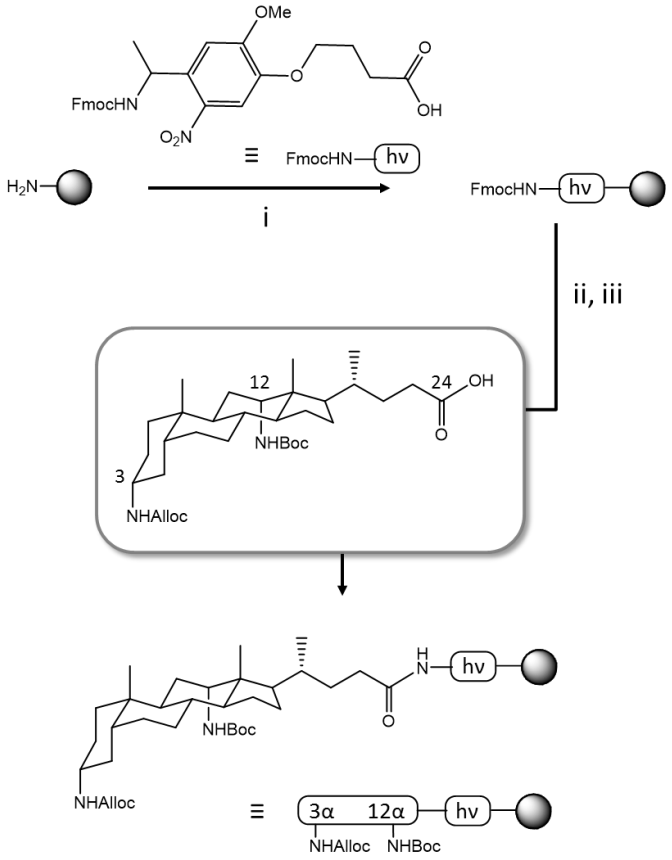


Figure 5.4: Synthesis scheme for the appendage of the linker and the scaffold to the solid support. i) PyBOP, DIPEA, DMF, 3h, rt. ii) 20% piperidine/DMF 2 x 30 min. iii) HATU, DIPEA, DMF, 2 x 4h.

Decoration of the C12 position with the first peptide should precede the C3 derivatization due to steric constraints imposed by the steroid structure and additional hindrance upon the presence of peptide at C3. Relying on the orthogonality of Alloc and Boc protecting groups, acidic deprotection afforded a free amine at the C12 position on which the protected amino acid residues corresponding to the cMyc or Max peptides were coupled. HATU/DIPEA couplings were performed for efficient peptide synthesis. Capping of the N-terminus was necessary for further decoration at the C3. Alloc deprotection of the amine at the C3 gave the second anchoring point for the coupling of cMyc or Max amino acid residues using HATU once again as an efficient coupling reagent for the sterically hindered amines. The final capping of the peptide strand at the N-terminus afforded the protected peptidosteroid tweezer. After cleavage from the resin and

deprotection of the side chains, the desired cMyc/Max bHLH TF model was obtained (figure 5.5) and subjected to DNA binding studies by the Electrophoretic Mobility Shift Assay (EMSA) (figure 5.6).

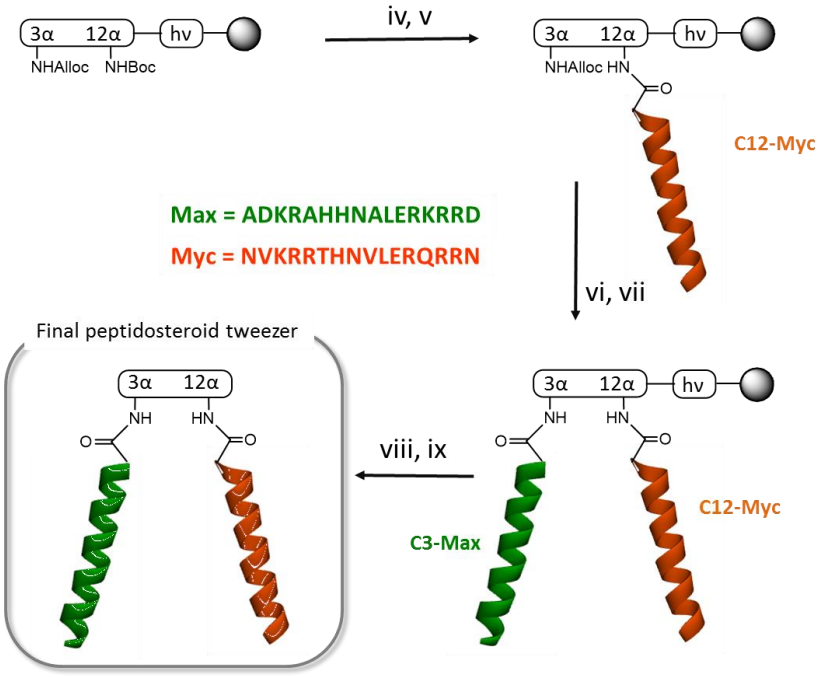


Figure 5.5: Synthesis scheme of model developed by Dieter Verzele. iv) 20%TFA/DCM 2h, rt. v) automated synthesizer. HBTU, DIPEA, 3h; 20% piperidine in DMF, 2, 5, 25 min. vi) Alloc deprotection. $Pd(PPh_3)_4$, Bu_3SnH, morpholine, DCM 2h rt. vii) automated synthesizer. HBTU, DIPEA, 3h; 20% piperidine in DMF, 2, 5, 25 min. viii) UV irradiation. ix) TFA/TIS/H_2O (95:2.5:2.5) 3h, rt, precipitation in cold ether.

Careful analysis of the results obtained from the EMSA indicated that the interactions were non-specific in nature as the peptide-DNA conjugate precipitated in the wells (Lanes 4 and 5). This was due aggregation, mainly based on electrostatic "sticking" of the mimics to the phosphate backbone, which is qualitatively very distinct from tight, sequence-specific insertion of α-helices in the DNA-major groove as observed with genuine TFs (Fig. 5.6).

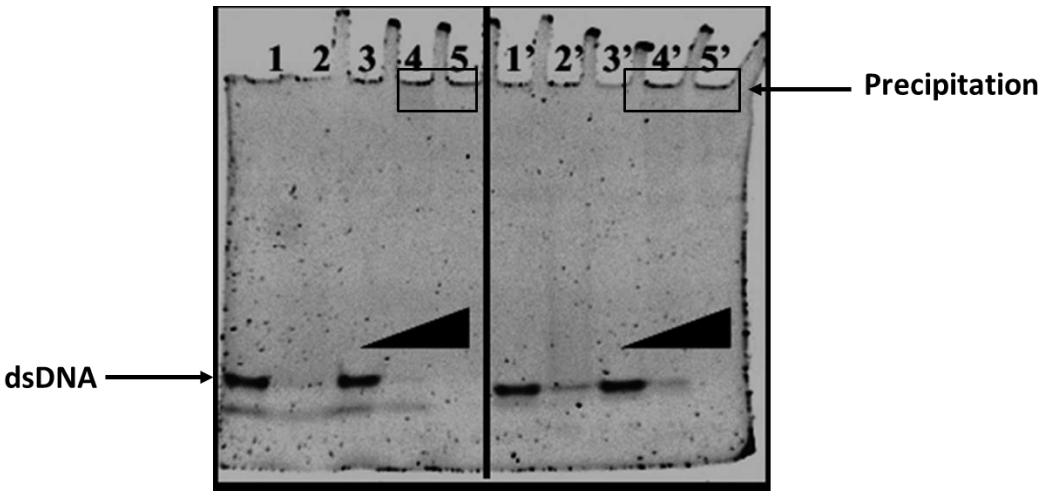


Figure 5.6: EMSA titration of the heterodimer and monomer (lanes 2 and 2') to the E-Box binding site (lanes 1-5) and to a random sequence (lanes 1'-5') to dsDNA (68.75 nM). Lane 1-5 contain respectively 0, 50, 4, 20, 40 µM, stained with SybrGold. Target dsDNA 5'-CTACTAGCACGTGCTAGTAG-3'. Random dsDNA 5'-GAAAATCACCCAACTGCA-3'.

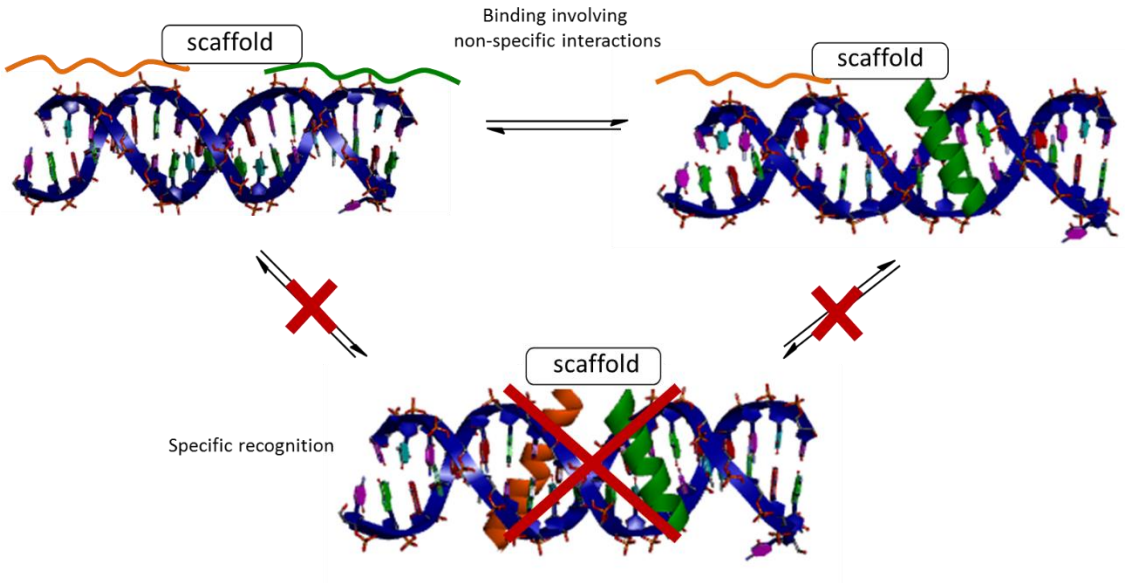


Figure 5.7: Pictorial illustration of different DNA binding modes to dimeric TF models.

Molecular visualization of the first generation design allowed the assessment of the structural restrictions for DNA binding at the major groove. Insight into the view of the cMyc-Max synthetic protein bound to its target E-box DNA sequence, determined at a high 1.9 Å resolution, provided an excellent model for present research[168]. This data can be found at the Protein Data Bank website under the name "PBD-ID: 1NKP". Visualization and evaluation of the DNA binder in the presence of the E-Box target sequence by modification of the PDB file features with Discovery Studio allowed the identification of the main problems for DNA recognition. The X-ray structure is shown in figure 5.8. The heterodimeric protein is depicted as helical ribbons with different color codes for each monomer.

As shown in figure 5.8, due to the direct appendage of the cMyc and Max basic region peptides onto the steroid scaffold, correct dimerization of these peptide helices was hampered. Because of the incorrect insertion of these peptide helices in the DNA major groove, essential contacts between the side chains and DNA are absent, causing the disruption of the DNA recognition. Due to the steric encumbrance of the peptide at the surface of the major groove, it was demonstrated that the model could not properly grip the DNA due to the rigidity imposed by the scaffold and the large space between the C-termini of both basic region peptides (Figure 5.6). Therefore, we proposed the need for an adapted design, once again indicating potential benefits of a spacer moiety between steroid scaffold and basic region peptides. This was to be done via the addition of 1 helical turn (approximately 4 amino acids) at the C-termini to the 17-residue cMyc/Max peptides. In following section, we will describe the design of an optimised Myc/Max bHLH TF model taking into account the synthetic conclusions from chapter 2.

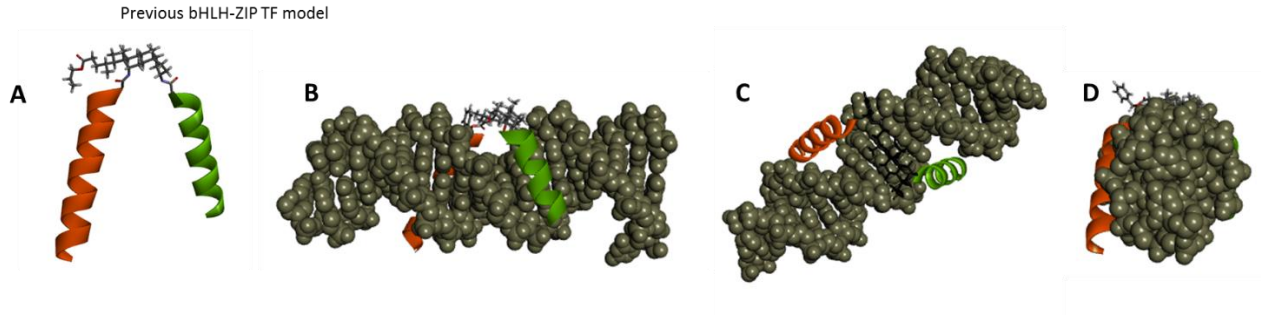


Figure 5.8: Molecular visualization of Myc/Max bHLH TF model (A) and its insertion into the DNA major groove from different perspectives (B, C and D).

5.3 Design of an artificial c-Myc/Max bHLH Leucine Zipper TF with the potential for sequence-selective DNA recognition.

Based on the conclusions of our previous results, we decided to elongate the basic region peptides at the C-terminus and incorporate a linker on the final peptide-steroid conjugate in order to increase flexibility and correct conformation in an attempt to achieve binding in the major groove of the DNA.

As a result of our successful methodology in chapter 2, a convergent approach was again applied for the appendage of the peptides onto the scaffold. In order to dimerize the peptides cMyc and Max onto the scaffold, an orthogonally-functionalized deoxycholic acid derivative was synthesized. This would then allow the incorporation of the basic region peptides of cMyc and Max oncoproteins in a sequential fashion. As we have previously reported, an aliphatic linker is needed for DNA binding of our transcription factor mimics. Therefore, we chose 4-maleimido-butyric acid and pentynoic acid as potential linkers to allow DNA recognition at the major groove and selective conjugation of both peptide strands onto the heterodipodal scaffold. The first conjugation at the C12 position was performed via a thiol-ene conjugation through a Michael nucleophilic addition to a cysteine-contained peptide resulting in a succinimidyl thioether moiety. The second conjugation was achieved via CuAAC at the C3 alkyne-functionalized scaffold. The conjugation was compatible with the presence of the peptide chain at C12. Both succinimidyl

thioether and triazole moieties are generally accepted as stable linkages under physiological conditions. Because of the results obtained when using CuAAC as a conjugation strategy for the GCN4 deprotected peptides to the 3,12-diyl bis(pent-4-ynoate) linkers of scaffold **2.29**, we considered CuAAC as the anchoring strategy for the challenging C3 position, sterically hindered by the presence of the monomeric peptide at C12.

The steroid nucleus afforded a unique advantage for the modification in case of the heterodimeric tweezers. The concave structure of the inner face of deoxycholic acid ensured different chemical reactivity for the alcohols at the C3 and C12 positions[173]. Esterification of the alcohol group at C3 is therefore favored over C12. Prior to this derivatization, a nucleophilic substitution with benzyl bromide for the protection of carboxylic acid at C24 was mandatory to avoid side reactions while performing the esterification of hydroxyl groups. Selective attachment of pentynoic acid to the C3 was achieved by Steglich esterification[174–176] in DCM by means of EDC (1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) and catalytic amounts of DMAP (4-dimethylaminopyridine) as an efficient catalyst that increases the reactivity of the carboxylic acid towards the alcohol group. Complete esterification at C12 with Boc-GABA-OH (4-(Boc-amino)butyric acid) was accomplished with EDC and stoichiometric amounts of DMAP in DCM. Acidic deprotection of the amine and subsequent maleimide formation provided the desired hetero-functionalized scaffold (figure 5.9).

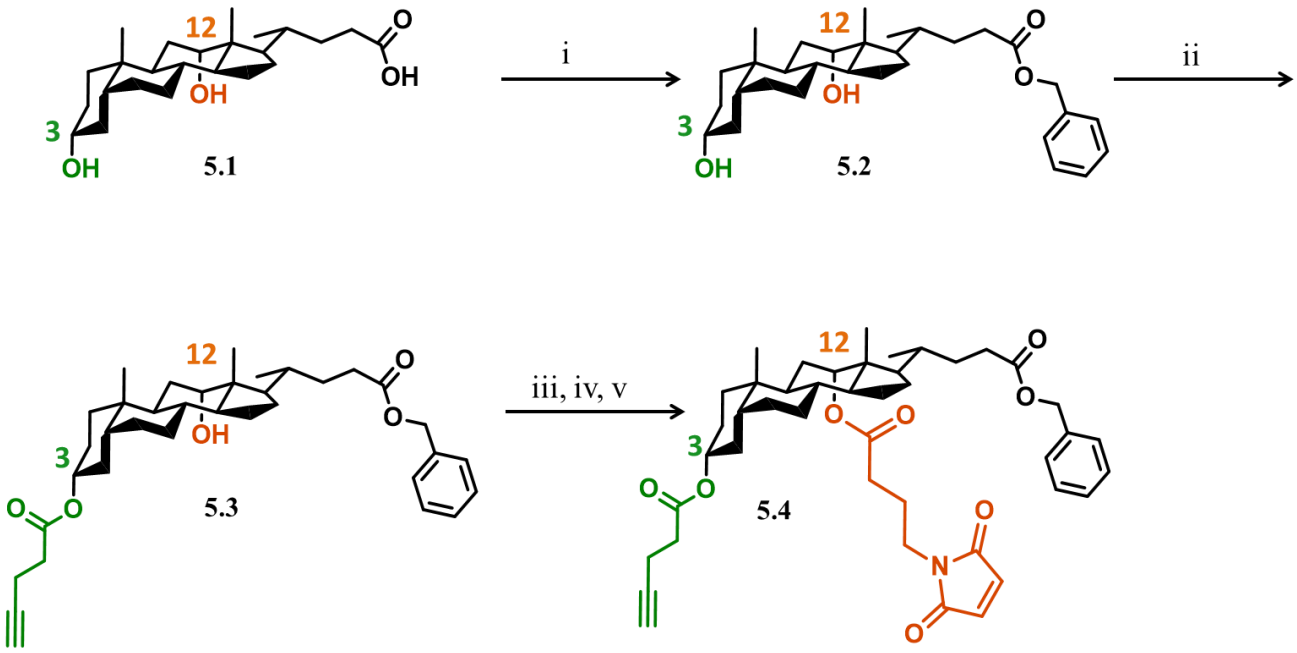


Figure 5.9: Synthesis scheme for the derivatization of deoxycholic acid (**5.1**) into the orthogonally functionalized deoxycholate derivative (**5.4**). i) DBU, Benzyl bromide, Dry DMF, 50°C, 24h. ii) Pentynoic acid, EDC·HCl, DMAP, dry DCM, rt, overnight. iii) Boc-Gaba-OH, EDC, DMAP, dry DCM rt, overnight. iv) 20% TFA/DCM, 2h, rt. v) Maleic anhydride, dry toluene. 80°C, 24h.

The elongated cMyc and Max basic region peptides were synthesized on a Rink amide ChemMatrix resin. Both peptides sequences were modified at the C-terminus. A cysteine was added to the C-terminus of the cMyc and the azido-homoalanine was added to the C-terminus of the Max. These residues were coupled manually. Further automated SPPS of the cMyc (NVKRRTHNVLERQRRNELKRGC) and Max (ADKRAHHNALERKRRDHIKDGAha) peptides was performed with HBTU as a coupling reagent and DIPEA as a base. N-terminal capping was done with 4-acetamido-benzoic acid and the final cleavage and deprotection using TFA/TIS/H_2O = 95/2.5/2.5 afforded the recognition units for the conjugation onto the scaffold.

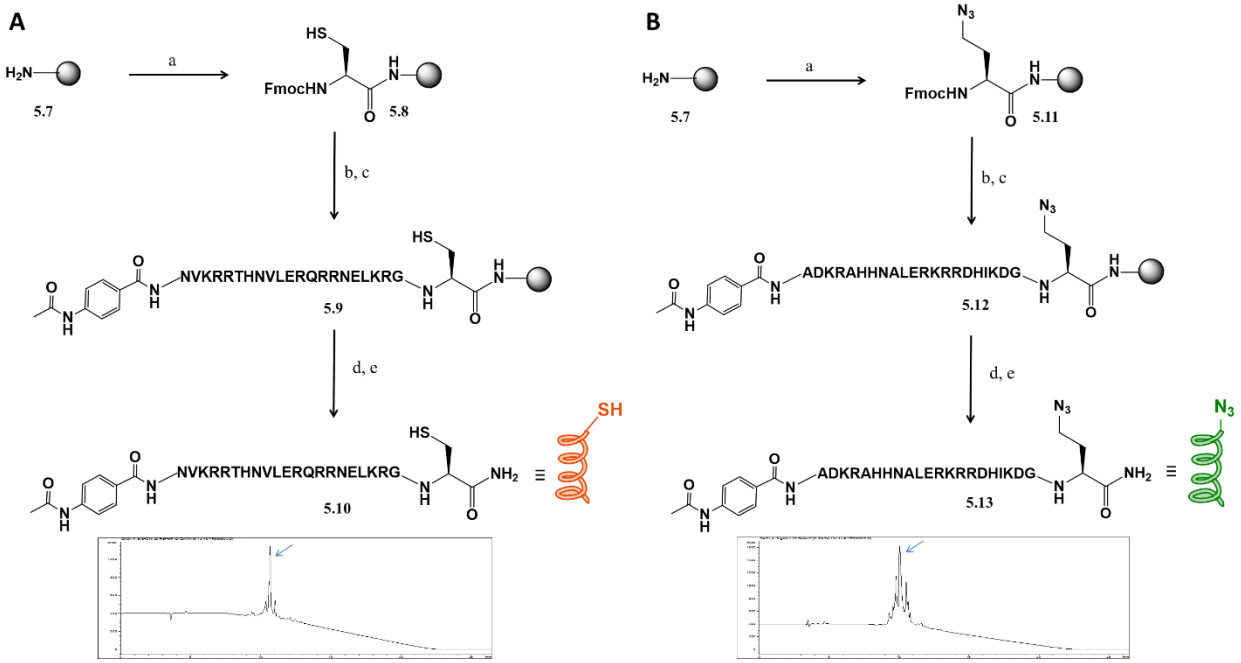


Figure 5.10: Synthesis scheme for (A) cysteine and (B) azido-functionalized GCN4 peptides (5.10 and 5.13 respectively) and crude RP-HPLC chromatograms (Jupiter C4 300Å, 0-100% ACN in 15 min). a) Fmoc-Cys(Trt)-OH or Fmoc-Aha-OH, PyBOP, DIPEA, Dry DMF, rt, 1h. b) 40% piperidine/DMF 2 x 30 min. c) Automated synthesizer. HBTU, DIPEA, DMF 1h. 20% piperidine/DMF 2, 5, 15 min. d) TFA/TIS/H_2O (95:2.5:2.5) 3h, rt. e) Precipitation in cold ether.

The heterodipodal peptidosteroid conjugate was synthesized by the conjugation of the cMyc-derived peptide to the maleimide-containing scaffold. The reaction was performed in ACN/H_2O (1:4) to ensure maximum solubility of both the peptide and the scaffold. NH_4HCO_3 buffer was used to fix the pH at 6.5. A slight excess of peptide was added. The reaction proceeded rapidly under mild conditions and with high specificity at room temperature and in an open atmosphere. Formation of the monomeric intermediate and consumption of the peptide was monitored by HPLC and MALDI. Maximum consumption of the scaffold was achieved after 2 hours. Presence of excess scaffold and absence of the peptide indicated that the impurities in the peptide in the form of salts or deletion sequences affected the ratio of peptide/scaffold. Because of the compatibility of the thiol-maleimide and CuAAC conjugations, we decided to continue with

the final step. Due to the compatibility of the excess reagents with the subsequent CuAAC, the crude reaction mixture was subjected to the conjugation at C3 without purification of the monomeric construct. Residue **5.5** was dissolved in DMSO and the elongated Max peptide dissolved in ACN/H_2O (1:4) was added in excess. $Cu(CH_3CN)_4PF_6$ dissolved in dry DMSO was then added to the reaction mixture. Consumption of the monomeric starting material was followed by HPLC. The reaction was found to be complete after 3h. Due to the excess of scaffold **5.4**, the monomeric peptidosteroid conjugate at C3 **5.6.b** was also present. Compound **5.6** was purified by RP-HPLC (figure 5.11).

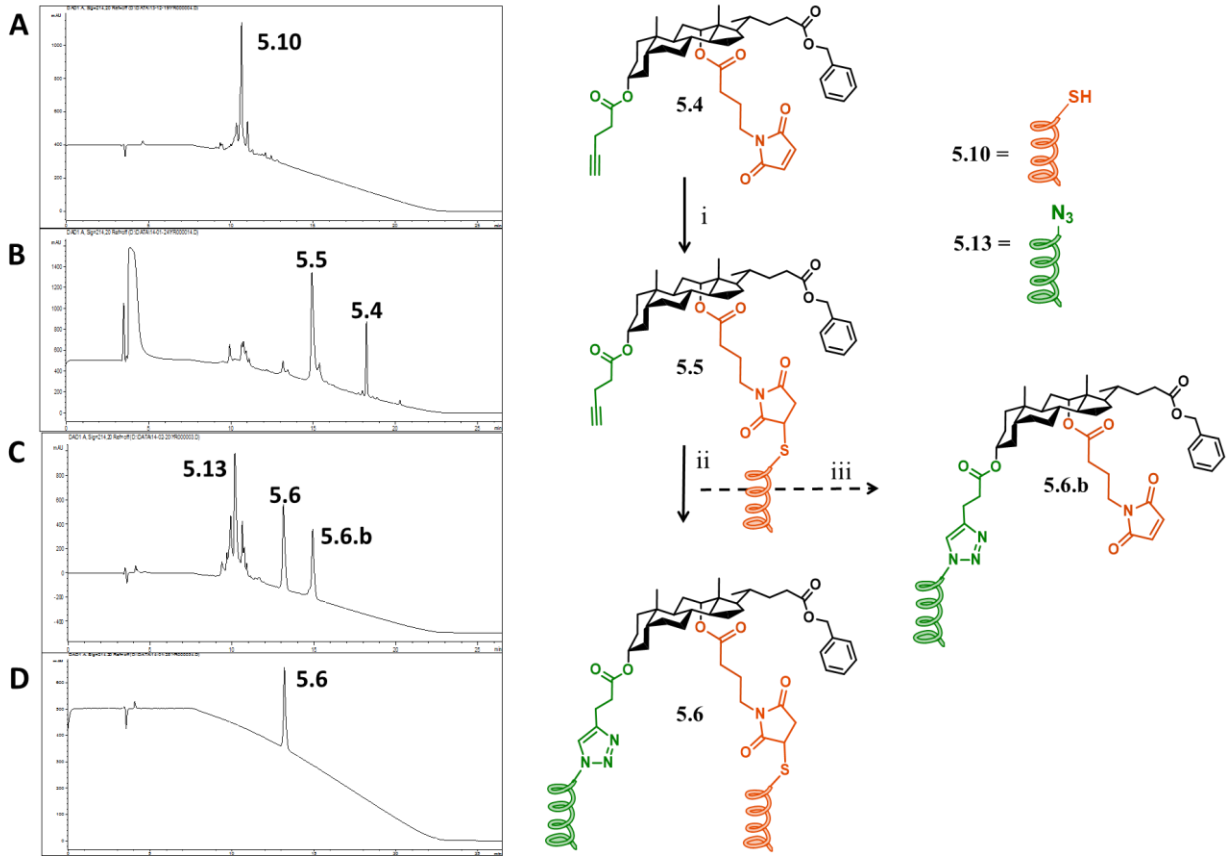


Figure 5.11: Synthesis scheme of final heterodipodal Myc/Max steroid conjugate **5.6** and RP-HPLC chromatograms of peptide **5.10** (A), cysteine conjugation reaction mixture towards **5.5** (B) and second conjugation via CuAAC to obtain final compound **5.6** (C), further purified by RP-HPLC (D) (Jupiter C4 300Å, 0-100% ACN in 15 min). i) H_2O/ACN 1:1 NH_4HCO_3 pH 6.5 rt, 2h. ii) $Cu(CH_3CN)_4PF_6$ DMSO/H_2O rt, 3h. iii) CuAAC to excess of scaffold **5.4** as side reaction.

5.4 Concentration determination via ERETIC-NMR

As reported in chapter 2 and 3, ^{1}H-ERETIC technique was again chosen as method to accurately determine the concentration of **5.6** for the EMSA study. Aromatic signals from the 4-acetamido-benzamide moieties at the N-termini, the 2 protons of each histidine of the peptides, the 5 protons of the benzyl ring at the C24 of the scaffold and the 5-H of the triazole were present at the aromatic region. The coupling pattern of the histidines is 4 singlets at 8.5 ppm from the 2-H and at 7.1 ppm from the 5-H of the imidazole moieties at

the side chains. The 5 protons of the benzyl ring at C24 show a multiplet at 7.25 ppm. The 4-acetamido benzoamide shows a more complex coupling than previous compounds, probably due to the fact that the last amino acids at the N-termini are different. Doublet of doublets are present at 7.65 ppm and at 7.42 ppm, each corresponding to four protons. On the other hand, the proton of the triazole is not visible, probably due to the low concentration of the sample and the low intensity of the signal (figure 3.8). The protons of histidine were used as signals to estimate the concentration of the sample from the reference sample. ^{1}H-NMR was performed in D_2O and the aromatic region was amplified for better integration due to the low concentration of the sample. The correlation between the integral values and the reference sample gave the concentration of the sample: 88 µM.

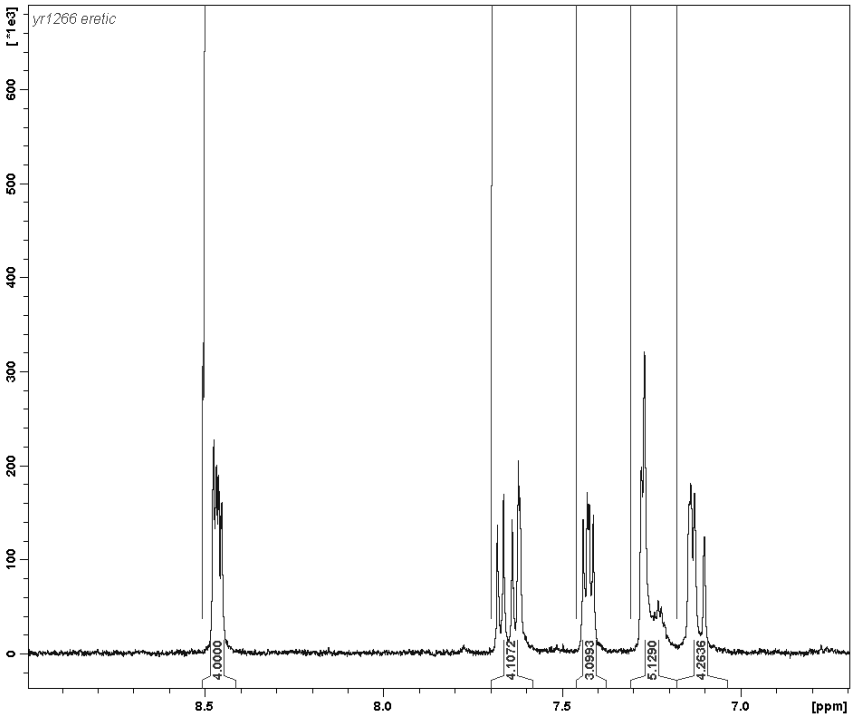


Figure 5.12. ^{1}H-NMR (D_2O, 500 MHz) of the aromatic region of compound **5.6** an integration of the signals corresponding to the 4-acetamido-benzoamide moieties at the N-termini of the peptides.

5.5 DNA binding studies using EMSA

The procedure for the EMSA study is described in section 10.4.

Figure 5.12 shows the formation of an upper band corresponding to the DNA-peptide complex formation. Due to the incomplete annealing of DNA, a ssDNA band on the bottom of the gel is present. Since there is no binding to the single strands, we hereby see the potential of the peptidosteroid tweezers to resemble the affinity of the natural bHLH TF families to their dsDNA sequence. In the last lane of the gel all material has disappeared probably due to unselective electrostatic interactions which are common at higher concentrations due to increase in both positive and negative charges and mass.

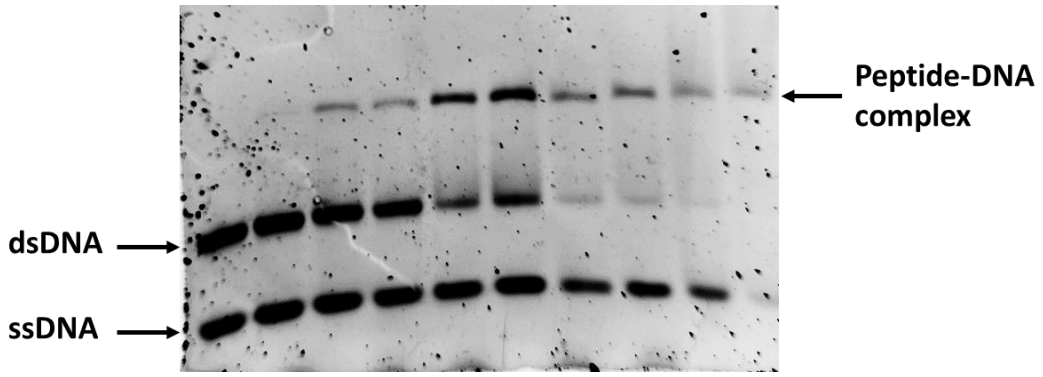


Figure 5.13: EMSA titration of the dipodal peptidosteroid to the E-box binding site stained with SybrGold. Concentrations for lanes 1-10 in μM: (0, 0.167, 0.334, 0.501, 0.584, 0.668, 0.751, 0.835, 1.169, 1.67).

For a better interpretation of the peptide-DNA complex, a molecular visualization process was performed for compound **5.6** with the dsDNA using Discovery Studio. In figure 5.13 it is observed that the dimer can perfectly interact with the major groove of the DNA.

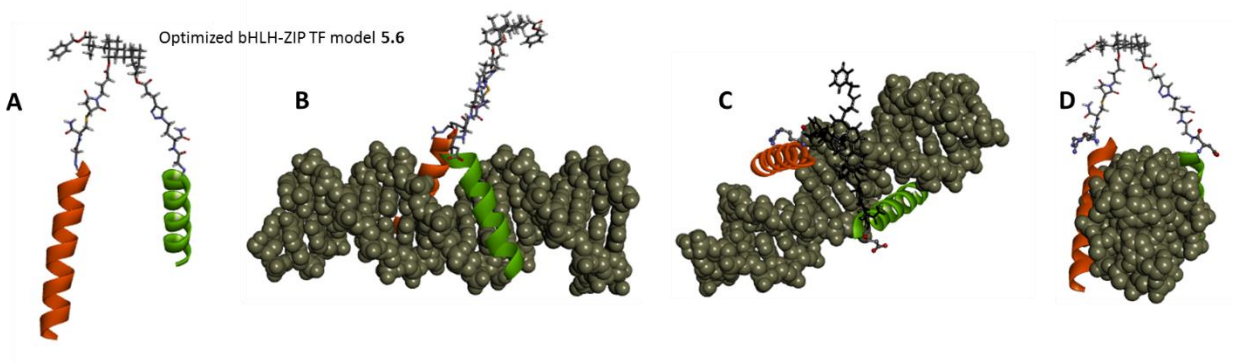


Figure 5.14: Molecular visualization of Myc/Max bHLH TF model **5.6** (A) and its insertion onto the DNA major groove from different perspectives (B, C and D).

In order to obtain a more quantitative picture, an electrophoretic Mobility Shift Assay (EMSA) with ^{32}P-radiollabeled DNA of the purified compound **5.6** was performed to test the DNA binding affinity of the synthetic mimic to the natural binding site of the protein (E-Box: 5' – CTA CTA GCA CGT GCT AGT AG – 3') in a suitable concentration range of the construct.

5.6 Conclusion

To conclude, we have presented here the first successful synthetic mimic of the cMyc-Max heterodimer. The potential of a reduced-size bHLH-ZIP cMyc/Max TF to recognize its dsDNA binding sequence was also tested and the results show that an optimized and miniaturized version of the protein contains all the necessary features to form a complex with dsDNA in a sequence-specific manner (Figure 5.12). In addition, we show the possibility of selective functionalization of steroid-based scaffolds to synthesize heterodimeric peptide-based conjugates, broadening the application of current methodology to other families of TF, such as bZIP Fos/Jun TF. In addition to the above, our design overcomes the difficulties of the decoration of

building blocks by the well-known SPPS, since it is based solely on conjugation strategies in solution phase. This can be considered the first model of an oncoprotein accessible by synthetic methods.

CHAPTER 6

DEVELOPMENT OF A NEW, CONVERGENT ONE-POT CROSS-RESIN LIGATION (OPCRL) APPROACH

Solid-phase synthesis to access peptidosteroid conjugates has been extensively researched by the OBCR group. The continuous exploration of this area has given the group new methodologies which overcome the difficulties of the linear coupling of amino acids to bile acid scaffolds. PhD theses prior to this one have been devoted to finding better synthetic routes for the same. The diverse application and broad scope of this unique conjugation has always encouraged the group to search for new conjugation methodologies. In this chapter we look at one such technique.

6.1 Introduction

6.1.1 Protein/peptide synthesis in solution phase

The synthesis of proteins has been a long-standing challenge. Isolation of proteins from complex biological mixtures is tedious and low yielding. Recent advancements in biotechnology utilizing site-directed mutagenesis and recombinant expressions have allowed synthesis of larger quantities of proteins, although with low yield due to aggregation issues. Despite the merits of these strategies, protein composition is limited by the ribosome-mediated biosynthesis. Protein structure, function and properties are further restricted by the genetic code. On the other hand, synthetic chemistry represents a source of versatility for protein modifications by segment ligations, incorporation of unnatural amino acids and building blocks and conjugation with other macromolecules. Development of new synthetic methods and improving the efficiency of peptide and protein assembly are thus one of the major challenges in bioorganic chemistry.

Ligation protocols for protein synthesis have been studied for more than 100 years, with the first precedent in 1906 developed by Emil Fisher[177]. Classical methods for ligation of fully-protected peptide segments consist of the condensation of the N- and C- termini in organic solvents due to solubility issues of the fragments. Due to poor purification, execution and analysis, ligation methodologies based on the conjugation of deprotected peptides by natural amide bonds were needed. Native chemical ligation (NCL) involving the chemoselective reaction of an N-terminal cysteine residue and a C-terminal thioester moiety allowed the exploration of natural amide bond formation between deprotected peptides.

Although NCL has been subjected to several improvements and achievements, synthesis of complex protein structures requires the condensation of more than 2 peptide fragments. Therefore, different orthogonal conjugation strategies involving several functional groups compatible with each other need to aim at

formation of natural amide bonds or the chemoselective formation of a non-peptidic bond. Within the latter one, bioconjugation strategies with wide utility involve of the formation of thioesters and thioethers. More recently the Copper Alkyne-Azide Cycloaddition (CuAAC) has gained particular attention due to the close resemblance of triazoles with amide bonds. On the other hand, amide bond formation can be achieved by other synthetic strategies that are compatible with NCL, allowing the synthesis of the backbone of native proteins. Next to the thioacid-azide amidation ligation and the reverse NCL, traceless Staudinger and KAHA (α-Ketoacid–Hydroxylamine) ligation are gaining particular attention.

The production of larger proteins requires the application of multiple ligation steps. The efficiency depends on the number of segments to be assembled and the yield of each ligation. Condensation steps mostly require repeated purification and lyophilisation procedures. Handling of peptide fragments results in low yields, hampering the generation of higher quantities of protein. In principle, a solid-support would prevent material losses and enhance the yield of the peptide segments assembly as in this case the purification and ligation steps can be avoided.

6.1.2 Protein/peptide synthesis on solid phase

Total chemical synthesis of proteins has been achieved with solid-phase peptide synthesis. It started however with peptide synthesis and not protein synthesis. The first description of solid-phase synthesis of a peptide was reported in 1963 by Merrifield[178]. He successfully attached an amino acid to the solid support via its carboxylic group, followed by the coupling of the next amino acid, bearing an activated carboxyl group and a protected amino terminus. After more than 30 years of improvements on the original methodology, the total chemical synthesis of a protein, the HIV-1 protease was achieved[179]. Despite the recent developments in solid phase chemistry by the introduction of new polymeric supports, linkers and protecting groups, the complexity of the protein structure and length hinder the access to protein synthesis by the sequential assembly of amino acids. On the other hand, complex proteins have been synthesized by the convergent assembly of protected or deprotected peptide segments by sequential chemoselective reactions on the polymer-supported polypeptide chain – the so called solid-phase chemical ligation (SPCL)[180].

Convergent approaches on solid phase towards peptide conjugations have not been limited to the synthesis of proteins. They are commonly employed for the preparation of bioconjugates of peptides and building blocks such as carbohydrates, cyclodextrins, bile acids and oligonucleotides. Bile acids have received special attention over the last decades due to a unique combination of rigidity and chirality coupled with a broad range of functionalization patterns with tunable reactivity, high availability and biocompatibility. Previous constructs developed in the OBCR group such as artificial estrogen receptors for endocrine disruptor chemicals[181], serine proteases[71] and haemagglutinin noose epitopes (HNE)[182] mimics and DNA binders[58,172] based on bile acid-peptide conjugates encouraged us to explore new synthetic strategies to overcome difficulties in the synthesis of peptidosteroid conjugate with long peptide chains mainly due to aggregation of protected peptides with the steroid. The limitations observed and encountered while using

peptide ligation in general intrigued us to investigate the development of a new convergent strategy for the synthesis of peptide conjugates.

6.2 New protocol for peptide chemical ligation

6.2.1 The need for a new strategy in peptide conjugation

Convergent ligation of peptides on solid-phase has been studied from a broad perspective and several difficulties have been encountered. Assembly of protected peptides to non-peptidic structures or to other protected peptides is synthetically challenging due to the poor solubility of such hydrophobic compounds. Losses while handling these segments and poor ionization makes the analysis of protected peptides complicated. Therefore, we felt a need to return to the basics of peptide handling and protection. From a practical perspective, we considered focussing on a solid-phase resin as a keystone for peptide synthesis. The sole precedent is of DeGrado and co-workers[183] who dealt with the synthesis of linear oligopeptides. The group of DeGrado employed oxime resin derived esters of a peptide, which was cleaved by transacylation with a nucleophilic activator to form an active ester in solution. The transfer of the peptide to the second resin occurred via a nucleophilic attack of a primary amine on solid phase, resulting in an amide bond formation. Due to the particular characteristics of this methodology, they named it as a Resin-to-Resin Transfer Reaction (RRTR). RRTRs are characterized as triphasic systems involving two independent solid-phases and a shared solution phase. It usually also involves a transfer between them and a reaction involving both resins as the name implies. Following this idea, the group of Hall reported the synthesis of arylglycine derivatives[184] and unsymmetrically functionalized biphenyl compounds[185] by RRTR. Consequently, we here present the possibility of applying the original RRTR methodology on the synthesis of peptide-based conjugates, an emerging area in supramolecular chemistry and multivalency[186–188] and of rejuvenating the concept of RRTR from its currently unknown status.

6.2.2 Description of our new protocol

First attempts towards the development of this protocol were carried out during the PhD thesis of Dieter Verzele. He designed the synthesis of a heterodimeric, peptidosteroid, tweezer-type construct by RRTR, providing a proof-of-concept to our new strategy. His initial results encouraged us to continue exploring the optimal conditions which could provide the final conjugate. We hereby describe the so-called resin-to-resin transfer reactions (RRTRs) to facilitate convergent peptide assembly by the one-pot conjugation of a resin-bound steroid and a polymer-supported protected peptide. The reaction is based on the in-situ liberation of the peptide from the donor resin and the appendage via amide bond formation to the acceptor resin bearing the steroid. This results in a peptidosteroid conjugate attached to a solid-phase (figure 6.1). Our new strategy requires the presence of two types of linkers, the side-chain protecting groups of the peptide and functionalities with the required reactivity for the ligation.

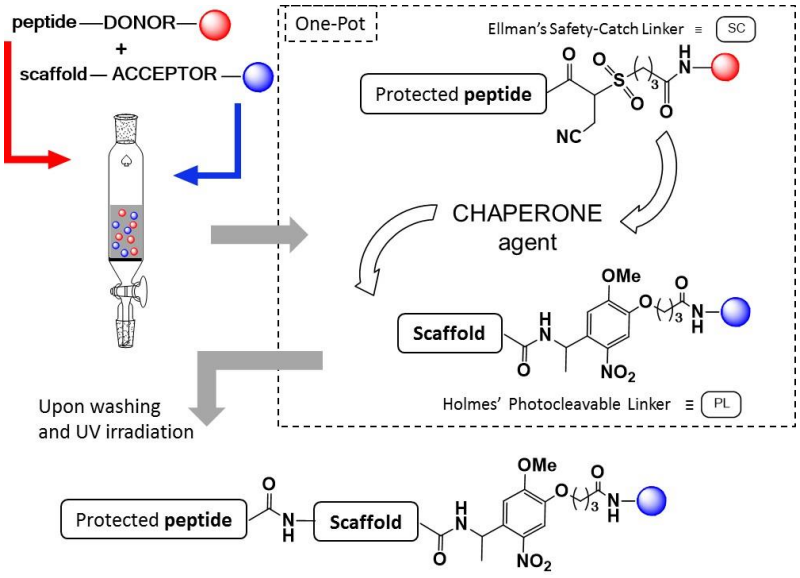


Figure 6.1. Triphasic Resin-to-Resin Transfer Reaction (RRTR) for the assembly of multivalent peptide conjugates, combining Ellman's Safety-Catch (SC) linker and Holmes' Photocleavable Linker (PL) with a chaperone that mediates solution-phase transfer by acting as both nucleophilic cleavage agent and activated carboxyl leaving group for in situ amidation.

Orthogonality is thus one of the main challenges of the present idea. RRTR requires the selective cleavage from one resin without affecting the other. Therefore both linkers should be orthogonal and the cleavage steps should not require acid treatment due to the undesired premature deprotection of the side-chains. Therefore, we have chosen a photolabile linker for the appendage of the steroid scaffold. Described by Holmes in 1995[88], the o-Nitrobenzyl photolabile linker allows the release of the final peptidosteroid conjugate by simple UV irradiation of the resin. On the other resin, the peptide is attached to the solid support via a safety catch linker. Pioneered by Kenner in 1999[189], he synthesized an arenesulfonamide linker and defined the safety-catch principle as "*a stable bond that is eventually labilised at the appropriate moment by a specific chemical modification*". The two-stage cleavage protocol is an essential feature of the safety-catch methodology. Release of the assembled peptide from the solid-support requires a distinct activation step prior to liberation. Intrinsic stability of the specifically-designed, inactivated linker moiety prevents premature loss of the compound. Further improvements of the proof-of-concept were developed by Ellman and co-workers[190,191]. They synthesized 3-carboxypropanesulfonamide as a safety-catch linker in which the activation involves the selective N-alkylation of the N-acylsulfonamide group. The cleavage is performed by the attack of a mild nucleophile (figure 6.2).

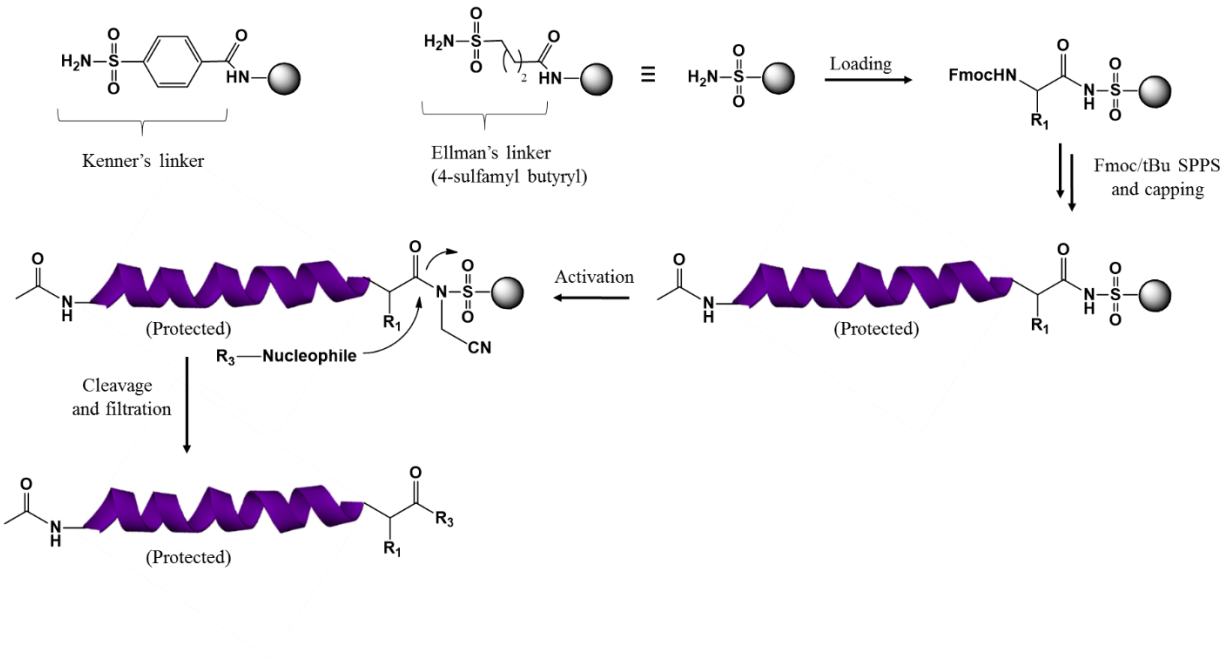


Figure 6.2: C-terminal modification of peptides by Ellman's safety-catch methodology.

6.2.3 The solid phase as keystone

The discovery of Solid-Phase Peptide Synthesis (SPPS) in 1963 by Merrifield[178] resulted in a turning point in peptide synthesis research. It facilitated faster and more efficient peptide purification, shortened reaction times and increased the final yield. Due to the mentioned advantages, SPPS was chosen as the preferred strategy for further improvements in peptide and protein production over the now obsolete solution phase peptide synthesis. Since the synthetic difficulty depends largely on sequence composition and properties, it encouraged the rapid development of a broad range of solid supports, side-chain protecting groups and coupling reagents. Regarding the bioconjugation of protected peptides on solid-phase, the success of the reaction relies on the structural, physical and chemical properties of the beads. In addition, the polymeric support should be mechanically stable, inert to a broad range of conditions and swell sufficiently in polar and non-polar solvents. Reaction progress depends on the grade of diffusion of the side-chain protected peptide in the resin matrix. RRTR requires two solid supports for the assembly of both the steroid and the peptide. During the reaction, the peptide should have easy access to the polymer-supported steroid scaffold. Therefore, the penetration of the peptide into the beads is an important factor controlling the progress of the reaction. Consequently, we proposed TentaGel resin as solid-support for both counterparts. TentaGel resin (Rapp Polymere Ltd.) consists of polyethylene glycol (PEG) units cografted onto a low-cross-linked polystyrene, which confers a hydrophilic character and good swelling properties in water as well as other polar organic solvents such as DMF and DCM. Studies on bead permeability have revealed that the molecular motion in the matrix of the beads is similar to that in solution. In particular, it has been shown that trypsin, a 23.5-kDa enzyme, can penetrate into the core of the 90-µm TentaGel beads[192]. Keeping these properties in mind, commercially available TentaGel-based solid supports were assessed and the TentaGel-NH_2 resin was chosen for the attachment of the scaffold. Preloaded H-Gly-Sulfamylbutyryl NovaSyn TG

resin was chosen for the appendage of the peptide. As discussed above, the sulfamylbutyryl safety-catch linker is used to ensure the cleavage of the peptide by nucleophilic substitution.

6.2.4 Synthetic procedure

The first step in the synthesis of the acceptor counterpart on TentaGel-NH_2 resin was the coupling of Holmes' photolabile linker using PyBOP/DIPEA as a coupling mixture. The lack of adequate monitoring via HPLC due to UV-inactivity of the scaffold was efficiently solved via the incorporation of the chromophoric moiety. Thus, the second residue to be appended was 4-aminomethyl benzoic acid, to follow-up the reaction progress by HPLC. Attachment of these moieties does not interfere with the subsequent RRTR methodology at the C12 (and C3) appendage. The orthogonally protected diamino deoxycholic acid derivative was coupled to the resin at the C24 position by a double-coupling with PyBOP. For straightforward decoration, the peptide generation at the C12-position should precede C3-derivatization in view of steric constraints. Already impeded by the steroid framework, additional hindrance upon prior C3-derivatization generally renders the C12-position inaccessible. Therefore it is mandatory that the reaction at C12 be done first. Boc deprotection afforded the free amine at C12 (figure 6.3).

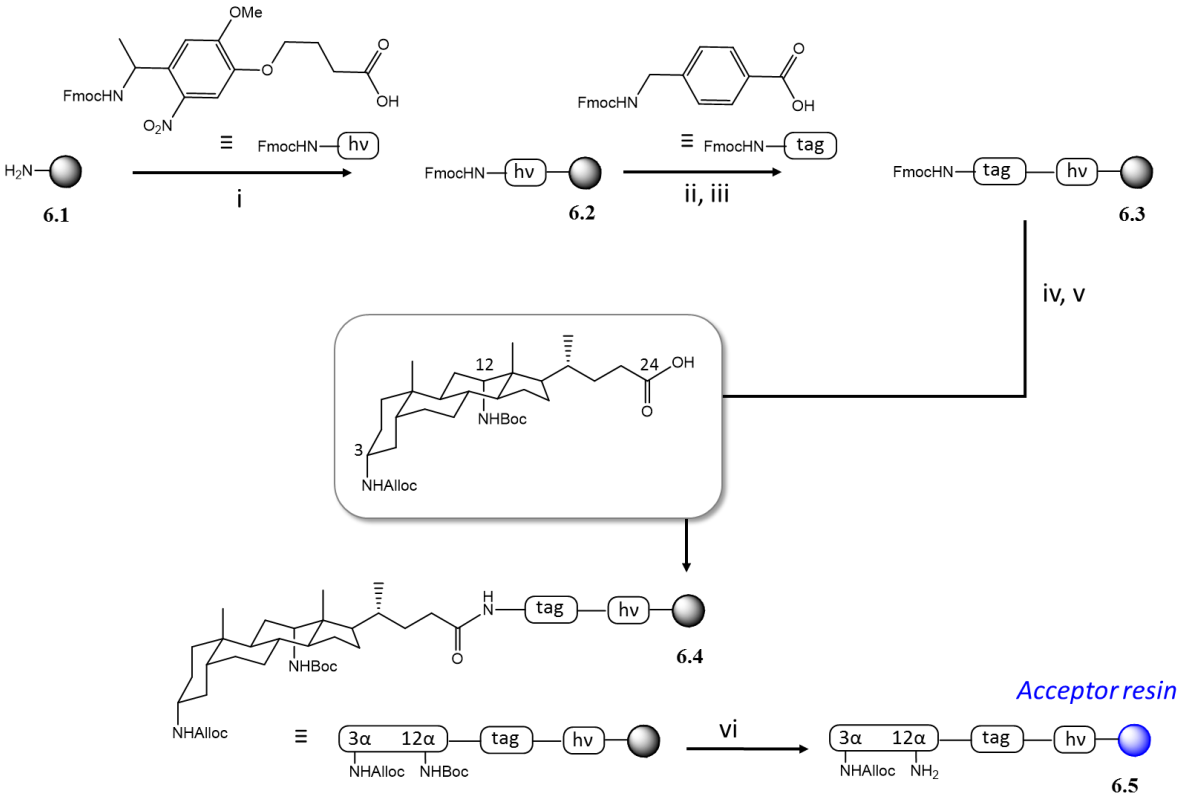


Figure 6.3: Synthesis route for acceptor counterpart. i) PyBOP, DIPEA, DMF, 3h, rt. ii) 20% piperidine/DMF 2 x 30 min. iii) PyBOP, DIPEA, DMF, 3h, rt. iv) 20% piperidine/DMF 2, 5, 15 min. v) HATU, DIPEA, DMF, 2 x 4h rt. vi) 20% TFA/DCM 2h, rt.

The peptide was synthesized on a preloaded H-Gly-Sulfamylbutyryl NovaSyn TG resin using standard Fmoc/tBu protocols in an automated peptide synthesizer using HBTU as a coupling reagent and DIPEA as a base. The peptide sequence comprised of the C-terminal residues of GCN4 basic region (A244-Q248). As both the unprotected or N-Fmoc protected amino group would interfere during nucleophilic cleavage of the peptide, an N-terminal capping is required to block the nucleophilicity of the amino-terminus. Capping was mediated through the treatment of the resin with 1-AcIm in $CHCl_3$, as a loss of compound has been reported when using acetic anhydride acetylation[193]. Prior to the RRTR, preactivation of the safety-catch linker was necessary to ensure proper liberation of the peptide during the reaction. N-cyanomethylation of the N-acylsulfonamide linker with iodoacetonitrile allowed cleavage of the amide bond by nucleophilic attack to the carbonyl group of the highly reactive N-cyanomethyl N-acylsulfonamide (figure 6.4).

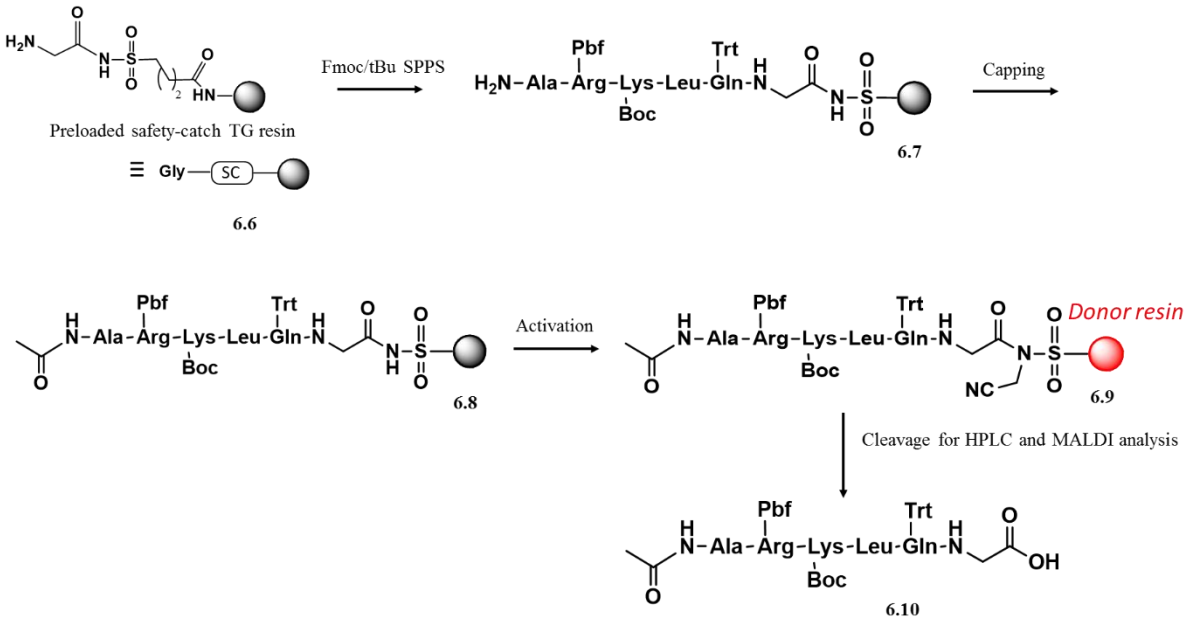


Figure 6.4. Synthesis route for donor peptide **6.9** comprising of the first 5 residues at the C-terminus of GCN4 basic region peptide.

Once the peptide was activated, the donor and acceptor resin were mixed in a peptide reactor. Incorporation of NMP as a solvent allowed the swelling of the beads. The reaction started with the cleavage of the peptide from the donor resin and the activation of the carbonyl group by reaction with a NaSPh/HOBt/DIPEA cocktail. The presence of a thiophenolate ensured the rapid cleavage of the peptide with a thioester functionality at the C-terminus. In-situ equilibration of the thioester group yielded the corresponding OBt-ester as a potential intermediate for facilitated coupling via the current convergent methodology (figure 6.5).

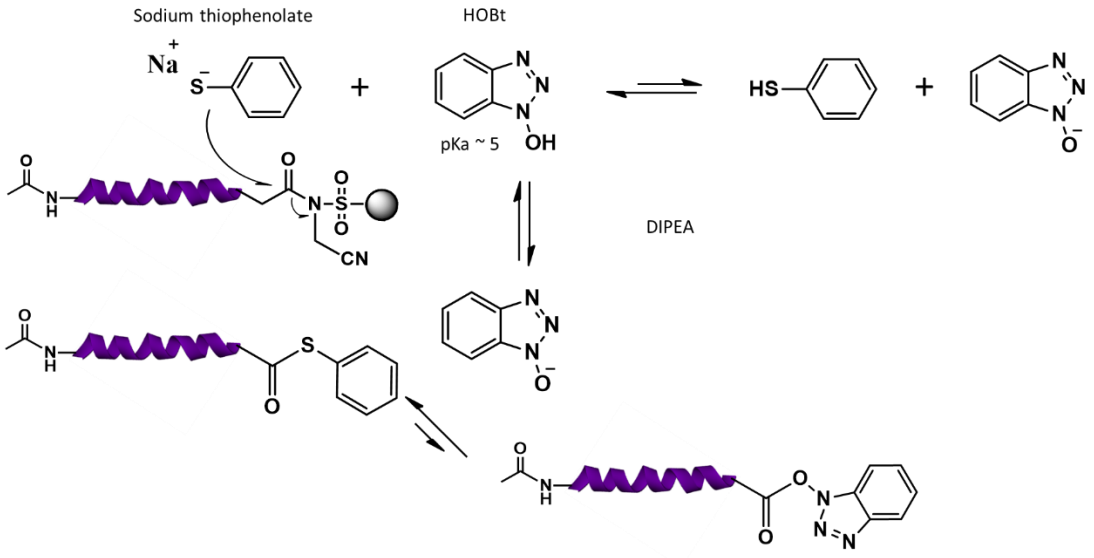


Figure 6.5. Tentative mechanism of cleavage and peptide activation.

A nucleophilic attack of the amine on the acceptor resin to the active OBt-ester resulted in the desired amide bond formation. A mere 5 equivalents of the donor-resin resulted in a significant conversion of starting material upon a 24h treatment. 50 equivalents of NaSPh, HOBt and DIPEA were added. Despite the basic properties of a sodium thiophenolate containing mixture, addition of DIPEA was essential for deprotonation of the accepting amine and thus ensuring its nucleophilicity. Washing and filtration steps allowed the removal of excess reagents and UV irradiation of a portion of the resin in ACN cleaved the C12-peptidosteroid conjugate from the acceptor resin (figure 6.6).

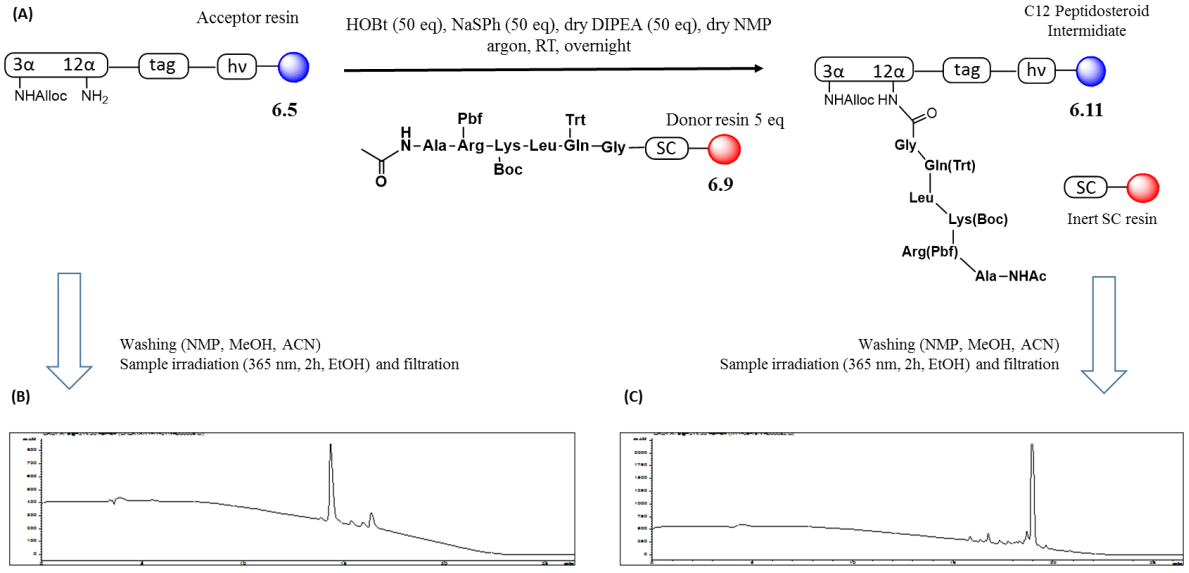


Figure 6.6. First RRTR to C12 bile acid position. (A) Synthetic conditions. (B) crude RPHPLC of starting acceptor after photocleavage. (C) crude RPHPLC after photocleavage showing complete conversion to (side-chain protected)

monomeric C12-peptidosteroid conjugate **6.11**. (cleavage as C24-carboxyamides, equivalents relative to acceptor resin) (RP-HPLC conditions: Jupiter C4 300 Å; gradient 0 to 100% ACN in 15 min).

To prevent hydrolysis of the activated ester moieties which could be detrimental for the product, carefully dried components were applied. Instead of the generally applied hydrate form, hydroxybenzotriazole (HOBt) was used as such.

Subsequent decoration at the C3 position towards obtention of the peptidosteroid tweezers by applying our RRTR methodology required the Alloc deprotection of the amine. The palladium-mediated hydrostannolysis with tributyltin hydride was used in the present work. Deprotection involved a palladium-catalyzed transfer of the allyl-unit to a nucleophile (i.e. a scavenger) in the presence of a proton source. The same reaction conditions for the RRTR at C12 were applied at C3. Unfortunately, no trace of the desired compound was observed after UV-irradiation. Therefore, further studies were performed for the optimization of the final conjugation via RRTR towards obtaining the peptidosteroid tweezers.

We concluded that there was a slow diffusion of the bulky, apolar side-chain, protected peptide intermediates in the cross-linked polystyrene matrix of the Tentagel-acceptor resin. Moreover, due to the sterical hindrance imposed by the scaffold, we proposed the incorporation of a linker at C3 to obtain a less hindered amine for nucleophilic attack onto the donor resin (figure 6.7).

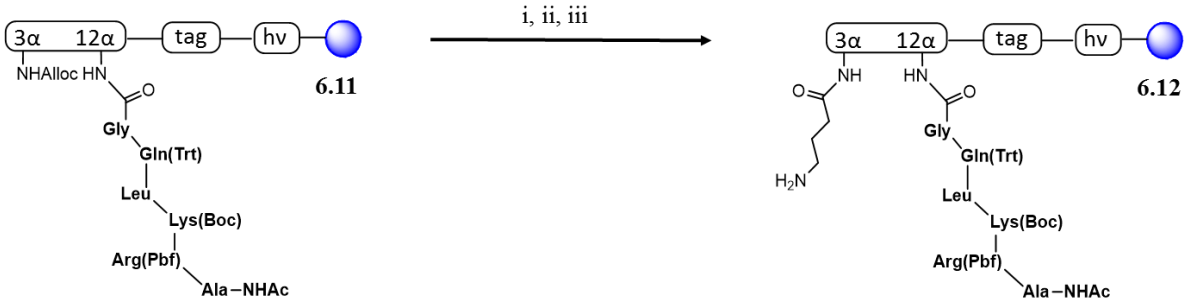


Figure 6.7. Derivatization of **6.11** with a GABA linker. i) Alloc deprotection. $Pd(PPh_3)_4$, Bu_3SnH, morpholine, DCM 2h rt. ii) Fmoc-GABA-OH, HATU, DIPEA, DMF, 4h rt. iii) 20% piperidine in DMF, 2, 5, 15 min.

A variety of chaperone agents were tested in order to find the optimum conditions for the conjugation at the more challenging C3 position: HOBt, HOAt, HOSu and Oxyma Pure (ethyl 2-cyano-2-(hydroxyimino) acetate)[194]. Again, in line with the DeGrado precedent, either HOBt or HOSu performed equally well. HOSu was a good choice considering commercial scarcity of explosive benzotriazole-based reagents. A similar excess of donor **6.9** was reasonable considering the resin accumulation during the reaction, which can inhibit the cleavage of the peptide from the resin and the posterior conjugation to the acceptor resin. A larger excess of DIPEA was also needed, due to the sterically hindered position of the amine at the C3 and the resin accumulation. Monitoring of the intermediate compounds was carried out by RP-HPLC. In addition, the protected peptidosteroid conjugates were analyzed on a Jupiter C4 300Å column for polarity reasons due to the hydrophobic nature of the compounds. After the deprotection of the side chains, the desired final compound could be analyzed in good integrity, with both RP-HPLC and MALDI (figure 6.8).

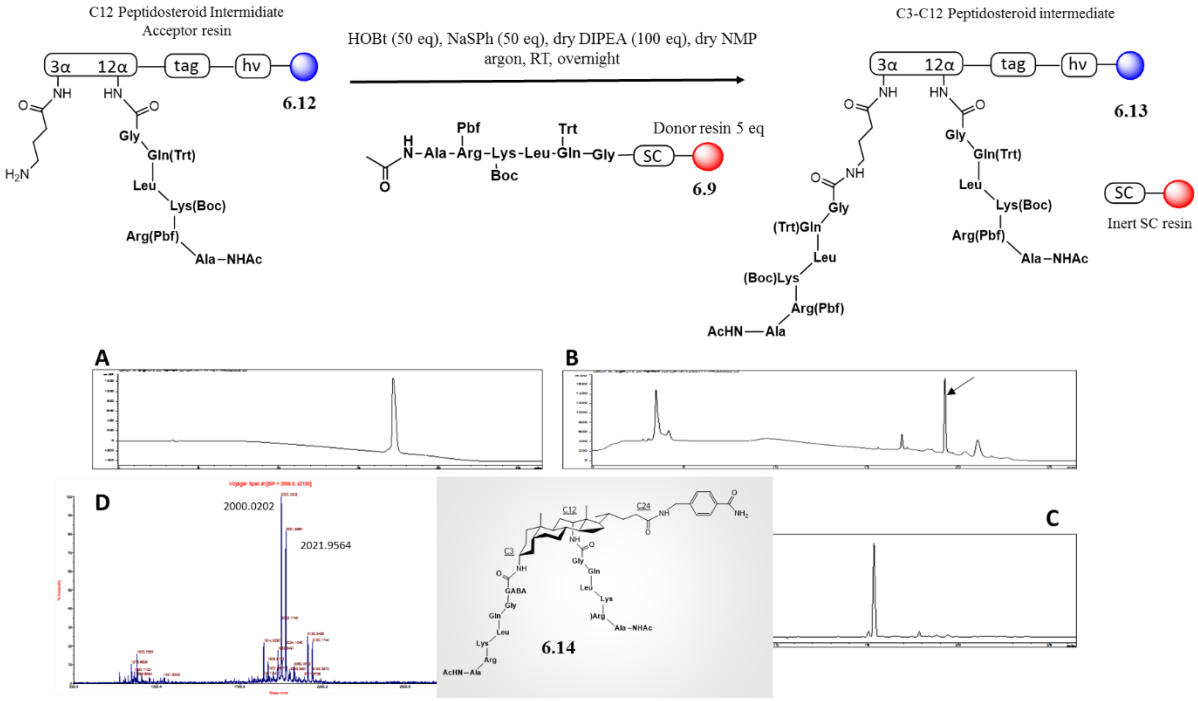


Figure 6.8. Second RRTR to C3-position upon Alloc deprotection and GABA introduction to yield the desired dimeric peptidosteroid upon final deprotection and cleavage. (A) Crude RP-HPLC of (side-chain protected) C3-GABA derived acceptor intermediate **6.12**. (B) Crude RPHPLC showing complete C3-RRTR conversion towards side-chain protected C12-C3 dimeric peptidosteroid **6.13**. (C) Crude RPHPLC and (D) MALDI of desired dimeric peptidosteroid upon final cleavage and side-chain deprotection **6.14**. (equivalents relative to acceptor resin) (RPHPLC with Jupiter C4 300 Å; 0-100% ACN in 15 min and MALDI recorded in the positive and reflection mode).

6.3 Conclusion

In summary, we hereby present an upgrade of the classical condensation strategy, without the practical encumbrances often associated with synthesizing, isolating/purifying and handling of the separate peptide fragments to be conjugated. We have achieved the synthesis of a supramolecular, multivalent (bio)conjugate with precise geometrical and appendage composition through a fully and genuine solid-supported methodology. Moreover, protected peptide fragment manipulations are troublesome due to solubility issues. Our convergent methodology combines the orthogonality of Ellman's safety-catch linker with the use of Holmes' photolinker to obtain multipolymeric schemes with notable advantages over the last Convergent Solid Phase Peptide Synthesis. Consequently, it is a shortcut towards synthesis of larger and more complex polypeptides and architectures. Furthermore, it compensates for time-consuming purification.

To highlight the main features of our procedure: there are no intermediates in solution which is typical for other ligations and excesses are easily removed. Therefore, handling of intermediate peptidosteroid conjugates is not necessary, ensuring higher yields and shortening the synthetic route. The proof-of-concept we present here is delivered using a scaffold featuring a combination of characteristics that include sterically hindered and concave positions. In close proximity, this could possibly cause problems such as incomplete

couplings with other ligation methodologies. Furthermore, we have achieved the strategic integration of Fmoc, Boc and Alloc chemistries, combined with the nucleophile susceptible Safety-Catch and photolabile cleavage.

As mentioned above, advanced orthogonality, a simple and fine-tuned procedure and an entirely solid-phase procedure characterizes our methodology. Moreover, our design can be adapted for other purposes such as to include an extra position between the scaffold attachment point and the photocleavable linker. This provides an extra dimension for multivalency which can be combined with further RRTR procedures as described herein. In conclusion, we describe the first use of resin-to-resin for multivalent decoration to access conjugates. Our strategy can be useful for solid-phase conjugations amenable to library formats. It is also applicable in areas as diverse as protein models, receptor enzyme design, chemical biology and classical medicinal chemistry, and compatible with microwave and flow chemistry.

CHAPTER 7

CASE STUDY: DESIGNING AN ARTIFICIAL SEQUENCE-SELECTIVE dsDNA NUCLEASE – SCOPE AND CHALLENGES

The current PhD thesis has been developed within a Marie Curie ITN Network, PhosChemRec (short for Phosphate Chemistry Recognition), in which different European universities and companies have collaborated with each other for the study of phosphate recognition and cleavage. Due to the expertise in phosphate cleavage within the group of Nick Williams at Sheffield University, his PhD students Zeyed Abdulkarim and Emmanuel Tirel were involved in a collaborative effort directed towards the obtention of sequence selective dsDNA cleavers. In this chapter we describe phosphate cleavage and the design of the catalytic system, based on our peptidesteroid conjugate as a DNA recognition unit.

7.1 DNA as storage for the genetic information

The human genome consists of an estimated 20000 to 25000 human protein-coding genes. Each relevant gene should ideally be transcribed into mRNA and then translated into a protein. Due to the important role of DNA as the carrier of the genetic information in the cell function, degradation or damage of the DNA sequences by mutations and loss of the mRNA can have dramatic consequences. Thus, the integrity of the genetic code depends on the bonds in which the nucleosides are linked together in both DNA and RNA. Nature has selected phosphodiester bonds as stable linkages for shielding of the oligonucleotides' composition and structure. The half-life of the hydrolysis of the DNA phosphodiester bonds at 25°C and pH 7.2 is approximately 31 million years, and has a $k = 7 \times 10^{-16}$ s^{-1} [195]. On the other hand, RNA is more prone to hydrolysis due to the presence of a 2'-hydroxyl group at the ribose ring. The estimated half-life for RNA is four years under the same conditions as those mentioned above for the hydrolysis of DNA[196].

This robustness of DNA is essential for the survival of all living beings from the tiniest simple cellular organisms to the one of the most complex being on earth in terms of genetic encoding – humans. If DNA was to degrade at a rate more rapidly than it was replenished in the human body, we would all have much shorter life spans. On the other hand, such hydrolytic inertness can also be a problem. Elimination of foreign DNA, excision and reparation of mutations and damages in the DNA sequence need to be addressed to ensure proper transcription and translation of the genetic information. To accomplish this task, enzymes called hydrolases are responsible for the hydrolysis of DNA at a physiologically relevant timescale. Despite the extreme inertness of the phosphodiester bonds, cleavage of these same bonds by natural enzymes is very efficient, providing impressive rate acceleration. As examples, serine or threonine phosphatases

accelerate the cleavage of phosphate monoesters by about 10^{21} -fold, while staphylococcal nuclease raises the rate of P–O cleavage in DNA by about 10^{17} -fold[197].

7.2 Cleavage of DNA: applications

The enormous scope of having access to artificial, site-specific DNA cleavage coupled with the recent progress in genomics, biotechnology and manipulation of DNA for the execution of several processes has encouraged researchers to look towards synthetic tools capable of cleaving nucleic acids. Molecular biologists have attempted to cut DNA at a site that is not recognized by natural restriction enzymes. This can only be achieved if synthetic constructs which recognize these DNA sequences and cleave them are available to us since cleavage at a non-recognized site is not possible using natural restriction enzymes. Recognition and incision of sequences within double stranded DNA (dsDNA) are widely used in gene manipulation[198] making this a highly pursued area of research. Thus, the design of synthetic enzymes tailored for the hydrolytic cleavage of DNA at the place of interest has found great utility[199].

In addition, synthetic hydrolases could serve as conformational probes for the determination of the three dimensional structure of nucleic acids, typically resolved by NMR and X-ray crystallography. Moreover, synthetic mimics could be helpful for the elucidation of the precise function of metal ions at the active site of natural hydrolases and also for the development of more efficient systems. DNA cleavers can also be found in antitumor antibiotics[200,201] and chemotherapeutic drugs[202–204]. In view of our advances on dsDNA recognition, we proposed the decoration of our DNA binders with cleaving agents for the sequence-selective cleavage of dsDNA.

7.3 Mechanism

Before we consider venturing into the challenging task of the artificial cleavage of phosphodiester bonds in DNA, a deep understanding of the mechanism of hydrolysis is required. Physical organic chemistry studies have provided a large amount of data concerning substitution reactions in phosphodiester groups[205–208]. In this chapter we explain the main points of the mechanism. The hydrolysis of phosphoester bonds involves substitution reactions and a pentacoordinate intermediate or transition state, called phosphorane and subsequent scission of the P-O3’ bond (scheme 1.1).

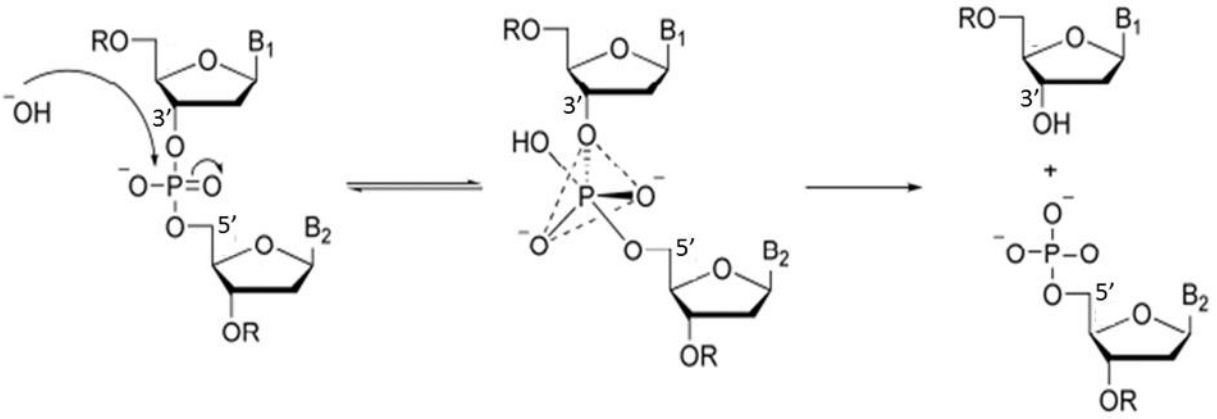


Figure 7.1: Proposed reaction mechanism for the hydrolysis of DNA. The enzymatically promoted 3'-O-P scission is shown. Lewis acids can be involved in activating water or the hydroxide as nucleophile or activating the phosphate group towards a nucleophilic attack (left).

Depending on the grade of substitution of the phosphoester, the reaction may follow three limiting mechanisms (figure 7.2)[197]:

1) The leaving group departure precedes the attack of the nucleophile and a metaphosphate intermediate is formed (DN+AN).

2) The attack of the nucleophile occurs before the departure of the leaving group and the formation of a phosphorane intermediate occurs. A two-step mechanism (AN+DN) is operative.

3) The bond formation between the nucleophile and the phosphorus atom along with the bond fission between the leaving group and the phosphorus occur simultaneously in a concerted mechanism (ANDN) and no intermediates are formed.

Phosphate esters in general follow the last mechanism, and the structure of the transition state depends on the esterification grade of the phosphate, moving from higher dissociative to more associative character as going from mono- to di- and to tri-esters[209].

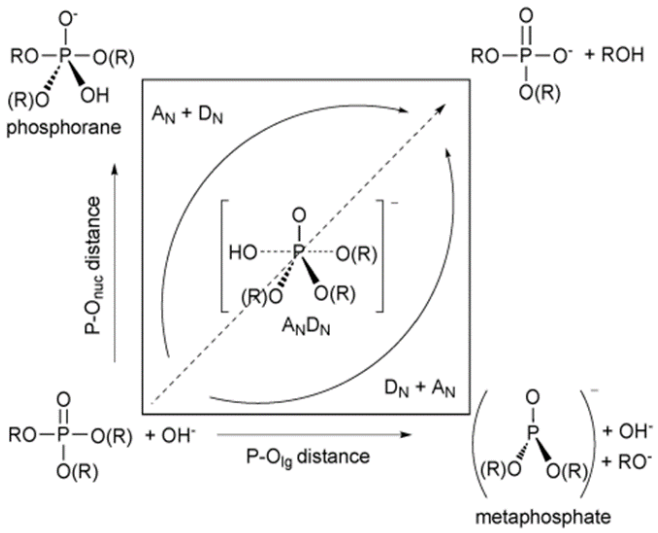


Figure 7.2. More O'Ferral-Jencks diagram for the mechanism of phosphate esters hydrolysis[4].

The common feature of these three mechanisms is the maintenance of a -1 charge on the phosphoryl group in the transition state[210]. Thus, in phosphomonoesters the leaving group possesses a full negative charge. In phosphodiesters, the phosphoryl group holds the -1 charge during the process. In the case of neutral phosphotriesters, the negative charge is given by the nucleophile in a tighter transition state. In addition, the reaction mechanism is affected by the pH. At lower pHs, protonation of the non-esterified oxygen atoms decreases the charge of the phosphoryl group, and in order to maintain -1 charge, the hydrolysis shifts towards a more associative mechanism.

Metal ions facilitate the hydrolysis of phosphoesters when acting as Lewis acids. The function of the metallic centers is the activation of the phosphate group for the nucleophilic attack, activation of water or hydroxide ions as nucleophiles or the increasing of the ability of the departing alcohol as the leaving group. These Lewis acids can activate the phosphate group towards a nucleophilic attack, activate water or a hydroxide as nucleophile or increase the leaving group ability of the departing alcohol[211]. The elucidation of the mechanism of hydrolysis of phosphate esters shows the properties that metals at the active site of enzymes possess. The metal binds the oxygen atoms and polarizes both the phosphate group and nucleophile. Therefore, both hardness and Lewis acid properties are desirable. To ensure turn-over of the catalyst, rapid exchange of the ligands coordinated to the metal should be feasible.

7.4 First attempts towards synthetic nucleases

In nature, hydrolysis of a phosphate bond can be carried out by different enzymes such as phosphatases[212], nucleases (including restriction enzymes)[213], and topoisomerases[214], all via similar mechanisms. A substantial amount of research has already been conducted to develop artificial nucleases for the hydrolysis of phosphate esters and the hydrolysis of DNA. However, due to the extremely complex nature of the

[4] Adapted from F. Mancin, P. Scrimin, P. Tecilla, *Chem. Commun. (Camb).* **2012**, *48*, 5545–59.

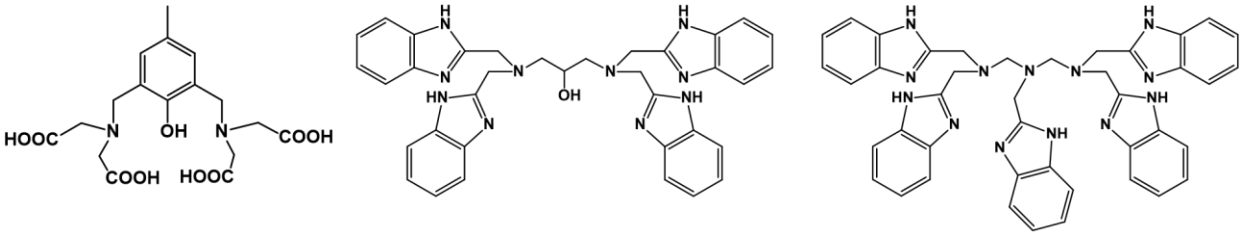
mechanisms and the kinetics involved there are still several questions unanswered. Some of the mimics which can be regarded as synthetic nucleases are described below:

7.4.1 The precedents: free ions and mononuclear complexes

In the early 90s, it was discovered that trivalent lanthanide ions accelerate the hydrolysis of DNA[215,216]. Ions do differ among themselves with regards to their affinity for DNA, but in general they show similar reactivity. Studies on Ce(IV) showed that this ion was 20 to 1000 times more efficient than any trivalent lanthanide[217,218]. Its higher efficacy was attributed to its very high electron withdrawing ability. Due to the toxicity of aqueous lanthanide ions in biological environments, it was concluded that free metal ions are not suitable for the design of artificial nucleases and that the formation of metal complexes is compulsory. On the other hand, formation of complexes with lanthanide metals decreases the activity in comparison with the free metal ion. One exception, however, is the Ce(IV)/EDTA complex which maintains the hydrolase activity. Other mononuclear complexes have also been reported, such as Co(III)[219] and Cu(II)[220].

7.4.2 Bimetallic complexes

Further improvements in design of artificial nucleases relied on the cooperation between two or more metallic centers to improve the catalytic efficiency. X-ray studies on natural metallonucleases with various metallic centers show that metal ions are in close proximity to each other. Consequently, an accurate location of the ions is needed when constructing the synthetic catalyst to benefit from the multiple interactions with the substrate and the activation modes that they provide to the hydrolysis. The first bimetallic centers for phosphate hydrolysis appeared in 1996, when simultaneous studies with mononuclear complexes revealed the potential of trivalent lanthanides to catalyze the reaction. Complexes containing two Eu(III), Pr(III) and Er(III) were reported, but low affinity for the substrate or low activity of the catalyst was obtained. In contrast to this result, double strand scission of plasmid DNA and linear dsDNA was achieved with Ce_2(HXTA)[221], with preference for the hydrolysis at the P-O3' site.



HXTA HPTB DPTB

Figure 7.3. Structures of HXTA, HPTB and DPTB ligands.

Besides lanthanides, Fe(III) is also present in phosphatases. Strategies towards bimetallic artificial nucleases based on Fe(III) have led to two main complexes, Fe_2(HPTB)[222,223] and Fe_2(DTPB)[224] (HPTB denotes N, N, N', N'-tetrakis(2-benzimidazolylmethyl)-2-hydroxy-1, 3-diaminopropane, and DTPB is the abbreviation of 1,1,4,7,7-pentakis(1H-benzimidazol-2-ylmethyl)-1,4,7-triazaheptane) (figure 7.3). The latter one has been thus far the best plasmid DNA cleavage agent under hydrolytic conditions. The case of

Cu(II) and Zn(II) is of particular interest[225], as bimetallic centers of these ions did not improve the efficiency compared to their monomeric counterparts. It was concluded that strict structural requirements in dinuclear complexes were essential and significant work had to be put in before success with Cu(II) and Zn(II) were achieved. It has been proposed that introduction of flexibility in the final constructs may overcome the impediments observed. In order to improve cooperativity between cleaving units in artificial models, it was proposed that these should be appended on the helical structure of a peptide by side-chain modifications and on the surface of nanoparticles, to promote the self-assembling. Peptides and nanoparticles afford cleaving moieties with sufficient flexibility to enforce a proper orientation which can allow phosphate cleavage.

7.4.3 DNA affinity units

Results on DNA cleavage by oxidative agents suggested that an increase in DNA affinity could lead to higher reactivity. The first examples reporting the employment of DNA affinity units onto synthetic catalyst were based on intercalators such as ruthenium and rhodium complexes, naphthalene units, phenanthridine and anthraquinone-based conjugates. More sophisticated models include the polyamides or Dervan peptides [226–228], PNA[229], peptides[230] and zinc-finger motifs[231].

7.4.4 Sequence selective elements

The major applications of DNA hydrolytic catalysts are the development of artificial restriction enzymes and chemotherapeutic agents. Since the natural enzymes are highly sequence-specific, the synthetic DNA cleaving units must also have this property as part of their structure. This could be achieved by the conjugation of cleaving moieties to DNA recognition units with specificity towards a particular sequence. Due to the convergence of two highly challenging aspects, phosphate hydrolysis and DNA recognition, few models have been built so far which can recognize dsDNA and cleave it at a specific site.

Since mimicking major groove transcription factors is synthetically challenging most of the models which can recognize a specific DNA sequence and cleave it use other DNA recognizers like the Dervan peptides (polyamides) or PNA or conjugated systems. Without going into too much detail a few examples are cited here. One of them achieved by Komiyama and co-workers[232], in which Ce(IV) complexated into a DNA-iminodiacetate cleaved a ssDNA at the P-O5’ site. Krämer followed another approach by the appendage of Zr(IV) complexes to PNA chains for the ssDNA cleavage at the P-O3’ bond[233]. Further improvements by Komiyama allowed dsDNA cleavage by Ce(IV)/EDTA complexes onto an oligonucleotide recognition unit[234] or complementary PNA chain[235]. Very recent models could achieve dsDNA cleavage randomly by a polyamide-Zn(II) complex conjugate[236] (figure 7.4 A) and specifically by a PNA-polyamide construct[237] through dsDNA invasion in addition to binding in the minor groove (figure 7.4 B and C).

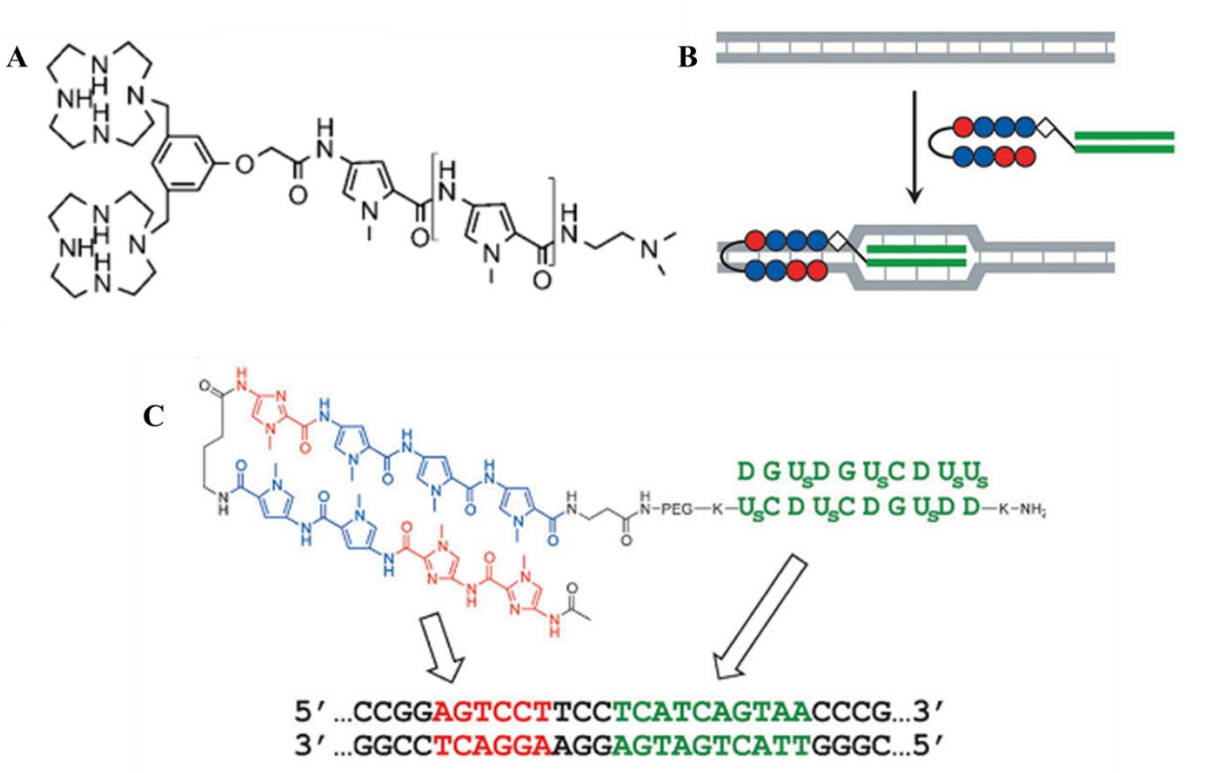


Figure 7.4: A) Structure of a polyamide-Zn(II)-based conjugate[236]. B) Scheme of the recognition of dsDNA by combining the PNA invasion and Py-Im polyamide binding (recognizing 6 bp DNA). C) Structure of the PNA/hairpin polyamide conjugate[237]. DNA sequences recognized by each portion are presented. PEG=polyethyleneglycol.

7.5 Our design: Towards a sequence-selective DNA nuclease inspired by the bZip TF

As reported in chapter 2, we have been inspired by the leucine zipper motifs (bZIP) for the construction of DNA binders. Due to the simplicity of the structure, in which the basic DNA recognition region binds to the major groove as a dimer, inserting two α-helices held in the correct position by a dimerization domain, the bZIP TF can be modified into a design which could allow incorporation of cleaving moieties for the sequence-selective cleavage of dsDNA in the major groove. Simplification of the synthetic route via a CuAAC between the basic region peptides and the steroid scaffold took us to the next step which was design of sequence selective DNA nuclease.

Encouraged by the low K_D values obtained, we envisioned the usage of our artificial DNA binder as an efficient construct which could be used to study the mechanism of phosphate transfer within a protein-like construct. Since this system can mimic the interaction of natural nucleases to DNA more precisely than synthetic metal complexes we believe it can provide excellent insight into dsDNA cleavage with artificial nucleases in the major groove. However, several challenges have to be overcome in order to achieve our goal. The high affinity to CRE must not be destroyed when the synthetic nucleases are incorporated. Also, proximity and scission at both strands of the DNA, sequence selectivity, cooperativity, flexibility, reactivity and turnover of the catalyst are some of the factors which need to be considered before starting any synthetic work. In view of the particular manner in which bZIP proteins bind DNA with each basic region binding

to only one DNA strand, we proposed the appendage of cleaving moieties in the side-chains of each peptide to achieve dsDNA cleavage. Despite the success in employing CuAAC for the conjugation, we foresaw possible competition between Cu(II) and Zn(II) ions for their complexation with the ligands. Therefore, we proposed maleimide-thiol conjugation for the appendage of basic region peptides to the scaffold (figure 7.5)

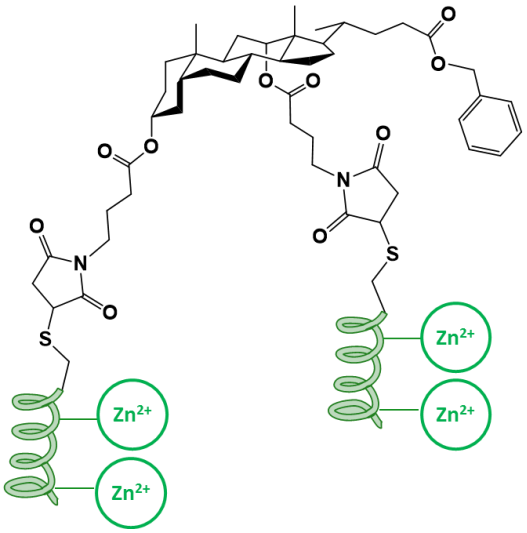


Figure 7.5. Design of DNA nuclease model based on bZIP TF.

7.5.1 Structural requirements.

As investigated in previous models, the models which enable cooperativity between bimetallic centers showed higher reactivity than those which employed mononuclear centers. The use of peptides as templates for the appendage of catalytic unit allows the incorporation of various metallic centers by modification of the required peptide side-chains. Moreover, the strict structural requirements to form a bimetallic catalyst are overcome by the flexible secondary structure of a peptide. Thus, cooperativity between metal complexes can be achieved due to the self-assembly of the peptides in a favorable conformation towards a phosphate group in a single DNA strand. As mentioned before the substitution of residues in the basic region peptides is challenging as it should not compromise DNA affinity. However, if non-DNA binding residues are modified the cleavers will not be oriented in the correct direction. Therefore, we reasoned that only those residues with no direct contact with DNA but having an indirect interaction (via a phosphate or water to base) can be modified. As GCN4 is an α-helix, the side chains of the amino acids located every 3.6 residues are oriented in the same direction. Since it is not possible to modify the 3.6th residue in peptide sequence modelling was performed on the i, i+3rd and i, i+4th residue to facilitate cooperation between the cleaving agents (figure 7.6).

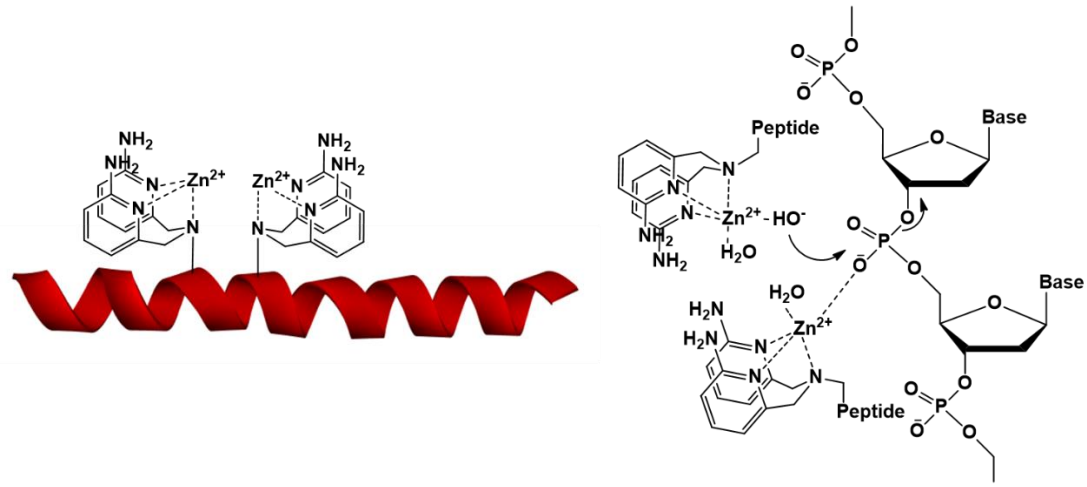


Figure 7.6. Bimetallic complexes on a peptide scaffold and their approach towards phosphodiester bonds in DNA.

As previously mentioned, DNA recognition should not be hampered. Therefore, a molecular visualization was performed on the system to select the residues that could be modified for DNA cleavage. As the amino acids responsible for the DNA interaction were mainly in the middle of the sequence, we proposed the alteration of the peptide at the C or N-terminus. By molecular visualization, we concluded that suitable amino acids are Ala244-Leu247 and Ala228-Lys231 (figure 7.7).

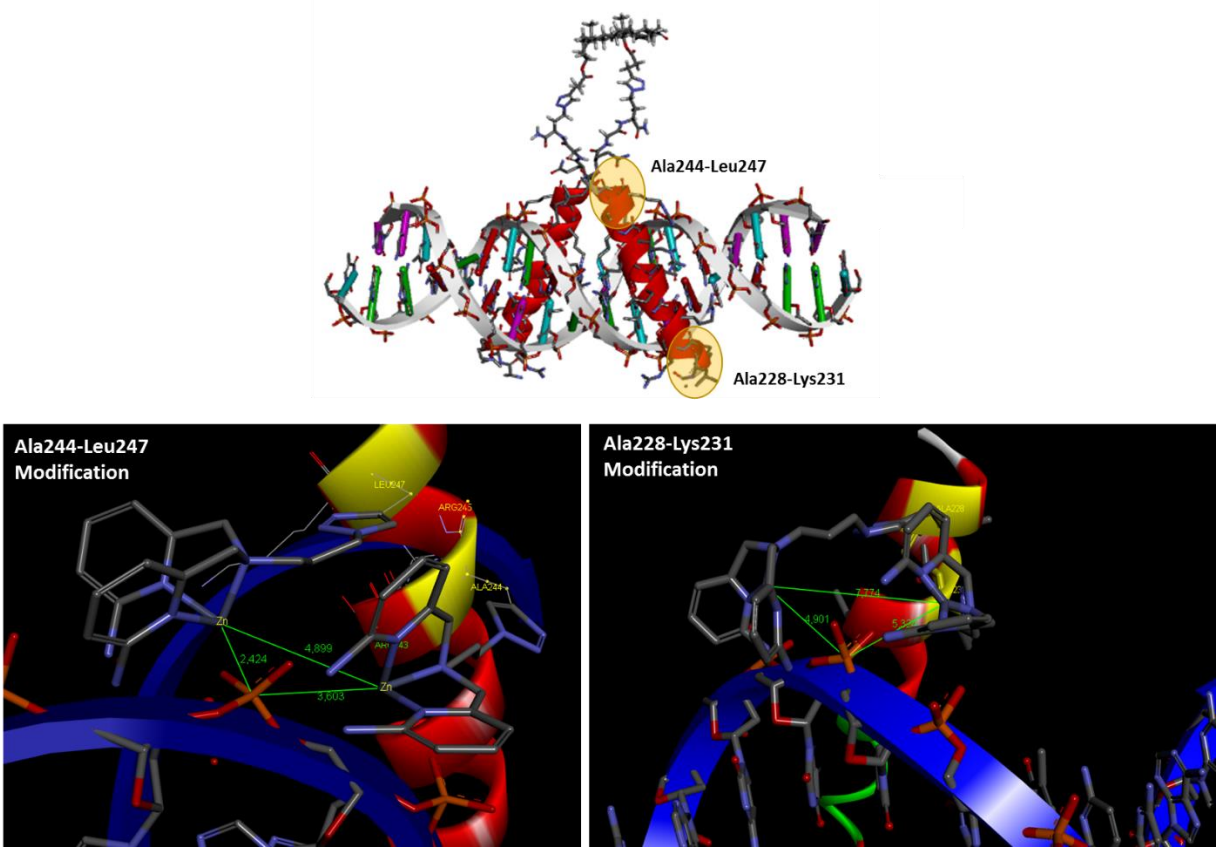


Figure 7.7. Molecular visualization of synthetic DNA nuclease and complex interaction with phosphates.

7.5.2 Synthesis of the scaffold

As mentioned in chapter 2, the steroid nucleus can serve as dimerizer for bZIP GCN4 transcription factor peptides. The optimization of the length and flexibility of the linker proved to be essential for enhancing sequence-specific DNA recognition. On the other hand, previous scaffolds were functionalized with alkynes or azides for the incorporation of the basic region peptides via CuAAC. Due to the possible interaction of copper in the cleaving ligands, DNA cleavage could be hampered as Zn ions will not be able to be complexed to form the cleaving units. Therefore, a different bioconjugation strategy was selected for the appendage of the peptides. Because of the availability of cysteine, and the possibility to functionalize the scaffold with maleimides, maleimide-thiol conjugation was chosen as method for the incorporation of GCN4 basic region peptides modified with the cleaving moieties.

The first step in the synthesis was the protection of the carboxylic acid at C24, followed by esterification of both hydroxyl groups at C3 and C12 with Boc protected GABA. Further deprotection of the amines and conversion to maleimides with maleic anhydride provided the desired scaffold for dsDNA cleavage (figure 7.8).

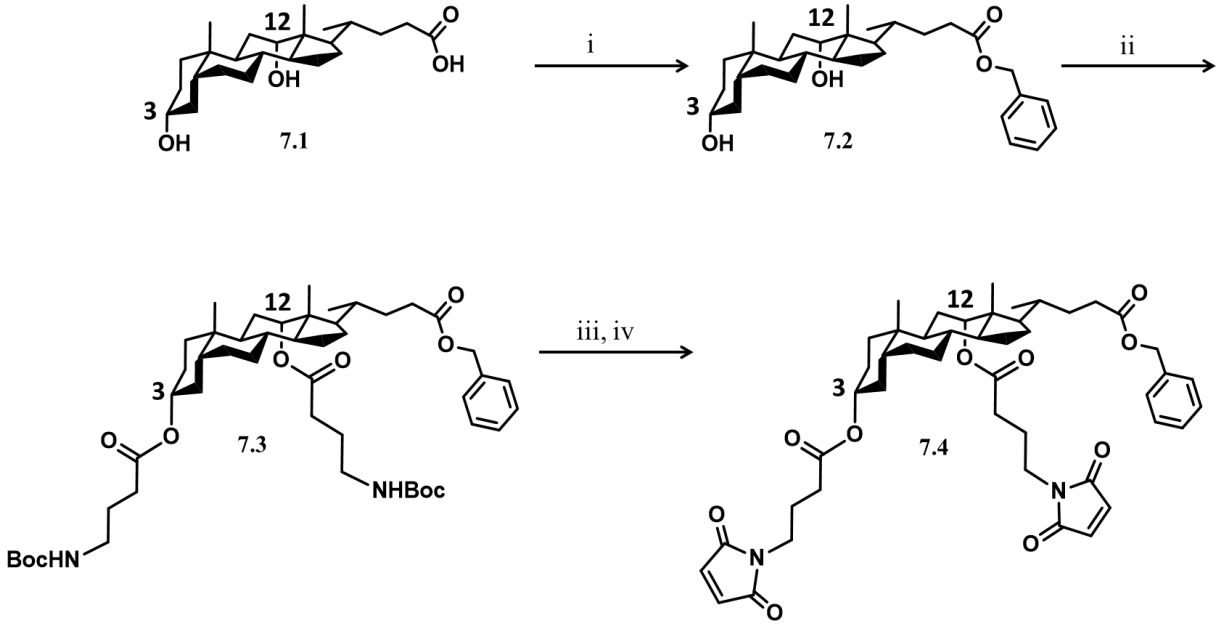


Figure 7.8: Synthesis scheme for the derivatization of deoxycholic acid (**7.1**) into the maleimide functionalized deoxycholate derivative (**7.4**). i) DBU, Benzyl bromide, Dry DMF, 50°C, 24h. ii) Boc-GABA-OH, DCC, DMAP, dry DCM, rt, overnight. iii) 20% TFA/DCM, 2h, rt. iv) Maleic anhydride, dry toluene. 80°C, 24h.

7.5.3 Cleaving units

Catalytic units were designed and synthesized by Zeyed Abdulkarim, a PhD student in the group of Prof. Nick Williams at Sheffield University. As a partner of the PhosChemRec Network, we had the opportunity to combine our expertise in DNA recognition with their knowledge of phosphate cleavage. In collaboration with Sheffield, we designed a construct (figure 7.5) and proposed a synthetic route for the incorporation of

artificial nucleases into our construct which showed sequence specific DNA recognition (figure 7.8). We envisaged this construct could be the first major groove dsDNA cleaver to be synthesized in a difficult albeit doable synthetic route. Previous studies by Nick Williams and co-workers[238–240] in Zn(II) complexes for phosphate ester cleavage served as a basis for the design of the cleaving moieties. The best results were obtained with complexes in which the ligands are based on pyridine moieties. In order to enhance the binding and the catalytic activity, two amino substituents have been incorporated to the pyridyl heterocycle as hydrogen bond donors for the stabilization of the transition state and to provide additional interactions in the active site[238]. Their successful results in bimetallic centers for phosphate esters hydrolysis prompted us to further investigate their incorporation on the basic region peptides. 2-aminobipyridyl ligands were modified for the assembly to amino acid side-chains via CuAAC (**7.6**). Formation of Zn(II) bimetallic centers is formed by the cooperation between monomers installed on residues A244 and L247 or A228 and K231.

During this project, we discussed which metal complexes should be included and the synthetic strategy for their incorporation on the peptide (figure 7.9). We envisage that excellent results obtained at the group of Prof. Nick Williams in the synthesis and study of rate-accelerating and catalytic turn-over Zn(II) complexes will allow further modification for its conjugation onto DNA recognition peptides. CuAAC conjugation towards compound **7.7** was achieved. Further improvements in the optimization of the purification step is necessary.

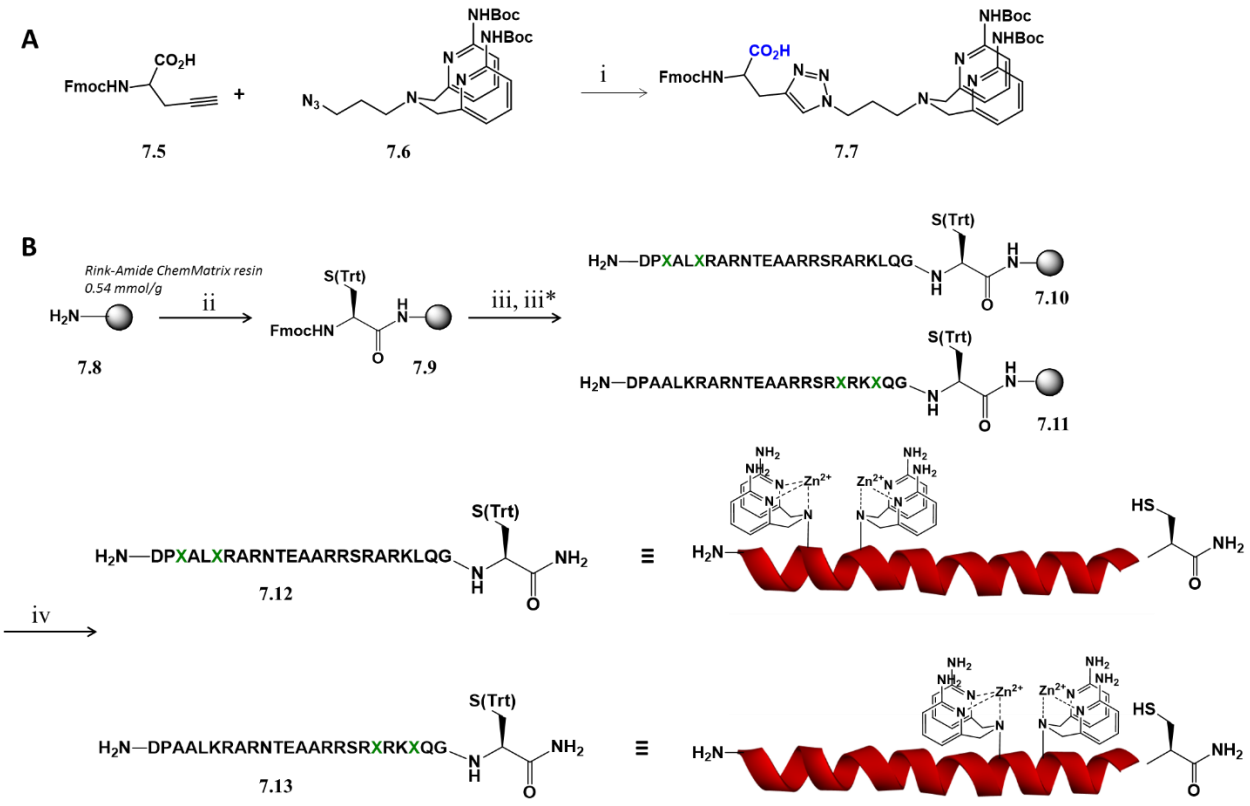


Figure 7.9. A) Synthesis route of cleaving moiety **7.7** with a free carboxylic acid (in blue) for further incorporation into the peptide by SPPS. Conjugation of Fmoc-Propargylglycine-OH (**7.5**) to cleaving ligand (**7.6**) by CuAAC (i) to afford cleaving moiety **7.3**. B) Synthesis route of GCN4 basic region peptides with cleaving moieties **7.7** to afford **7.12** and **7.13** peptides. ii) 1st coupling: Fmoc-Cys(Trt)-OH 0.5M; PyBOP 0.5M; DIPEA 2M in Dry DMF, 2h, rt. Capping: AcO, DIPEA 2M, dry DMF; 2x30min. iii) Automatic synthesizer: cycles AA corresponding to GCN4 0.5M; HBTU 0.5M; DIPEA 2M in NMP, 1h, rt. Fmoc deprotection: 20% piperidine/NMP; 2, 5, 15 min. iii*) Manual coupling of cleaving moiety **7.7**. Cleaving moiety 0.5M; HATU 0.5M; DIPEA 2M in Dry DMF, 4h, rt. Fmoc deprotection: 20% piperidine/NMP; 2, 5, 15 min. iv) Cleavage and deprotection: TFA:TIS:H_2O (95:2.5:2.5) 3h, rt followed by precipitation in cold ether.

7.6 Conclusion

In this chapter we present our initial idea for the synthesis of artificial dsDNA nucleases on the major groove, based on transcription factor mimics. We have synthesized the required scaffold for the appendage of the peptides by maleimide-thiol coupling and we have identified the positions at which the cleaving moieties can be incorporated on the peptide for the optimal approach to the phosphates of the dsDNA. On the other hand, optimization of the catalytic activity of the ligands and the synthesis of the final catalytic moieties for the incorporation on the peptides need to be performed.

CHAPTER 8

COLLABORATIVE PROJECTS WITHIN UGENT

8.1 Design and synthesis of an artificial estrogen receptor as endocrine disruptor chemical binders.

The first months of my PhD were dedicated towards the continuation of the work of Sara Figaroli, a former PhD student who defended her PhD in 2010[181,241,242]. The main aim of her research was the development of synthetic receptors for the binding of endocrine disruptor chemicals (EDCs). In this section, the synthesis of a new type of synthetic receptor for endocrine disruptor chemicals designed by Sara Figaroli is presented.

EDCs are environmental pollutants and possess estrogenic activity. A large interest around these compounds has been generated as they interfere with hormone production and regulation and the reproductive systems of humans and wildlife. EDCs cover a broad range of compounds – from synthetic and natural hormones to pesticides and additives used in industry for the production of plastics. In addition, they are active at concentrations even below 1 ng/L. Thus, trace determination of EDCs in environmental samples is vital because of their activity at extremely low concentrations. Unfortunately, the trace determination of their presence in environmental samples is also tedious at such low concentrations given the limited tools at our disposal. Moreover, the analysis of EDCs in water is not straightforward due to the complexity of the environmental matrix. Examples of EDCs and their impact when binding the Estrogen Receptor at the Hormone-Binding Domain (ER-HBD) have been determined for diethylstilbestrol (DES), bisphenol-A (BPA), and 17β-estradiol (E2) (figure 8.1) among others[243].

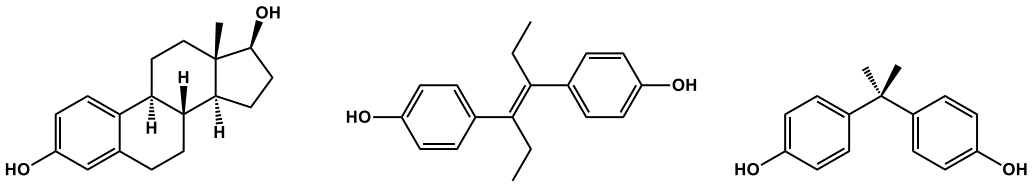


Figure 8.1. EDCs studied via X-ray diffraction of their corresponding ER-complexes. From left to right: Estradiol (E2), diethylstilbestrol (DES) and bisphenol-A (BPA).

To solve the challenging problems described above, the development of a high-throughput system was proposed which involved the pre-concentration of EDCs for further analysis. It was based on a solid-phase extraction (SPE) cartridge in which the artificial receptor was appended on a polymeric surface.

The design of the artificial receptor was based on the structural data of the ligand-binding cavity of the Estrogen Receptor (ERLBC) with its natural ligand (E2) available at the Protein Data Bank (PDB). The binding pocket is constituted by amino acids that interact by hydrophobic and hydrophilic interactions with

the ligand. Thus, studying these interactions was crucial for the rational design of the synthetic receptor. The detailed analysis of the data of PDB 1ERE (figure 8.2)[244] enabled the characterization of the single interactions between each amino acid with 17β-estradiol (E2). It was shown that the phenolic hydroxyl group of the A-ring interacted via hydrogen bonds with the carboxylate of Glu353, the guanidinium group of Arg 394, and a water molecule. The alcohol functionality at C17 established a hydrogen bond with Hys524. The remain parts of the E2 were involved in hydrophobic contacts via its A, B and D rings with the side chains of Ala350, Leu387, Met388, Leu391, Phe404, His524 and Leu525.

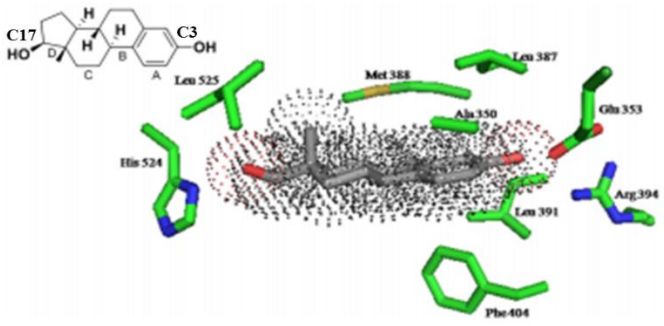


Figure 8.2. Detailed view of ER-LBD complexed with E2 (1ERE). Only amino acid side chains with contact residues are depicted.

In order to obtain a three dimensional cavity and better mimic the natural receptor, a steroid scaffold was used as the skeletal structure for the receptor. For a better mimic of the cavity, a cage-like receptor was designed by linking scaffolds together[245,246]. Further decoration of the scaffold with the necessary amino acids should provide the hydrophilic contacts to E2. With the aim of obtaining the optimal interactions between the amino acids, the hydrophobic moieties and the E2, we proposed a steroidal cage-like structure. This structure could afford the required hydrophobicity, while hydrophilic contacts were maintained by the replacement of the imidazole unit of histidine by a triazole ring and the arginine and glutamic acid were positioned at the C24 on the steroid nuclei, resembling the face-to-face interaction in the natural binding cavity. Due to the essential presence of the triazole, the cage-like structure could be obtained by CuAAC of both steroidal parts (figure 8.3).

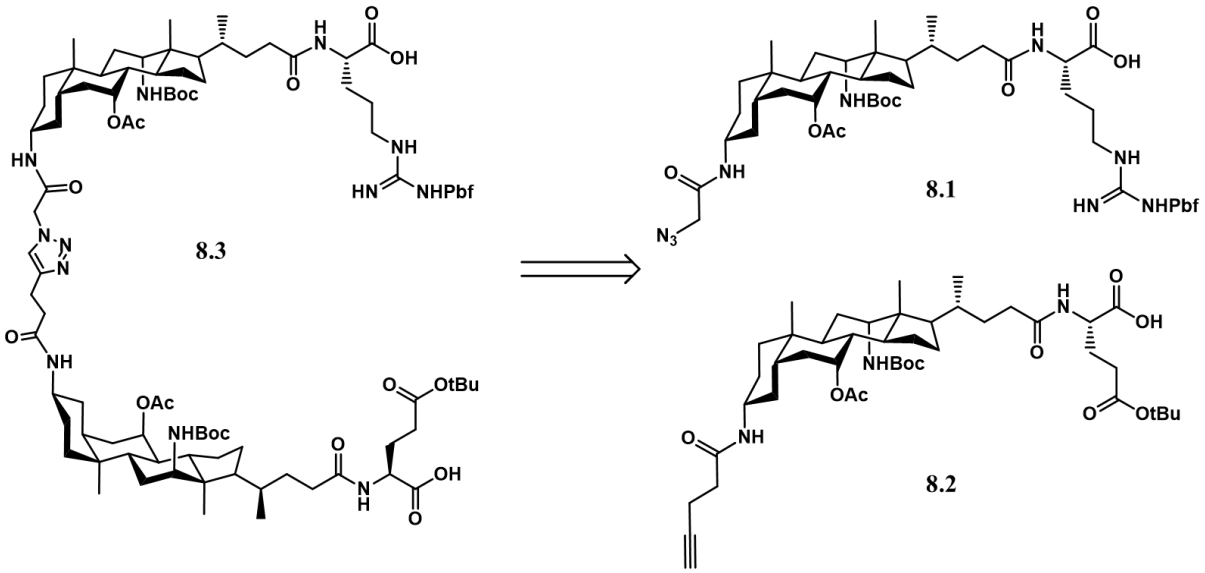


Figure 8.3. Structure of the cage-like receptor **8.3** and its retrosynthesis.

The synthetic receptor in figure 8.3 was modelled by Dr. P. Hendrickx using the software Accelrys Discovery Studio 1.7 (figure 8.4a). From those results, we could conclude that the cavity should allow accommodation of a molecule of E2. An overlap of the cage-like receptor in blue with the main residues of the natural ER-HBD in green complexed to E2 is depicted in figure 8.4b. The main structural differences between both receptors is: 1) the distance between the side chains of glutamine and arginine and 2) the position of both amino acids relative to the E2 molecule, as they are not facing the phenolic hydroxyl group in the artificial receptor.

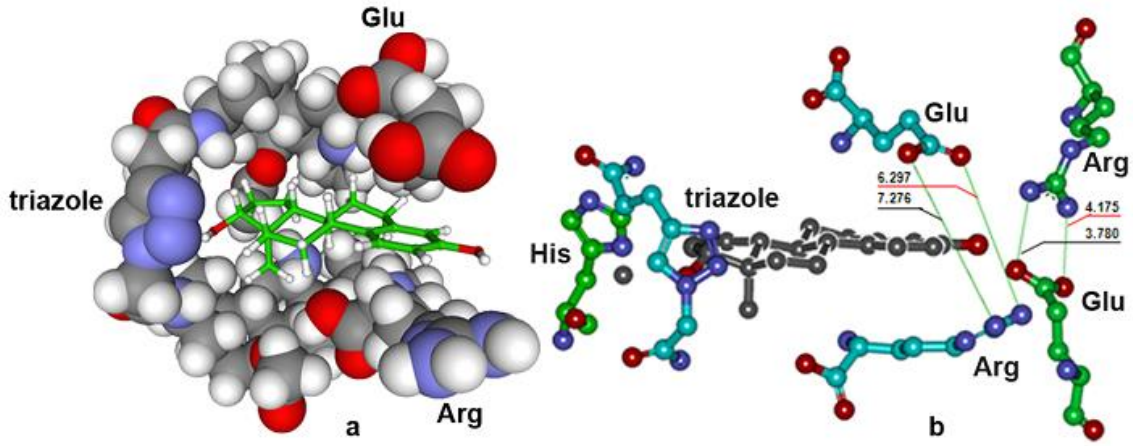


Figure 8.4. Molecular modelling of receptor with E2 (a) and comparison with the natural receptor (b)[1].

[1] Adapter from: PhD Thesis of Sara Figaroli. "Development of a Synthetic Receptor for Trace Determination of Endocrine Disruptor Chemicals". Promoter: Prof. Dr. Annemieke Madder. Ghent University. 2012

For the evaluation of the binding capability of the synthetic receptor to E2, diffusion NMR was performed. Unfortunately, it was shown that the receptor did not have affinity for E2. As an improvement to this design for future studies, we suggested that a spacer between the amino acids arginine and glutamine and the scaffold may help to accommodate a molecule of 17β-estradiol.

8.1.1. Synthesis

The synthesis of both alkyne and azide-containing components **8.1** and **8.2** were performed on a 2-chlorotrityl chloride resin. The first step was the coupling of the amino acids arginine or glutamine followed by the anchoring of the scaffold. Further Alloc deprotection provided a free amine for the appendage of the linkers with an alkyne or an azide (figure 8.5).

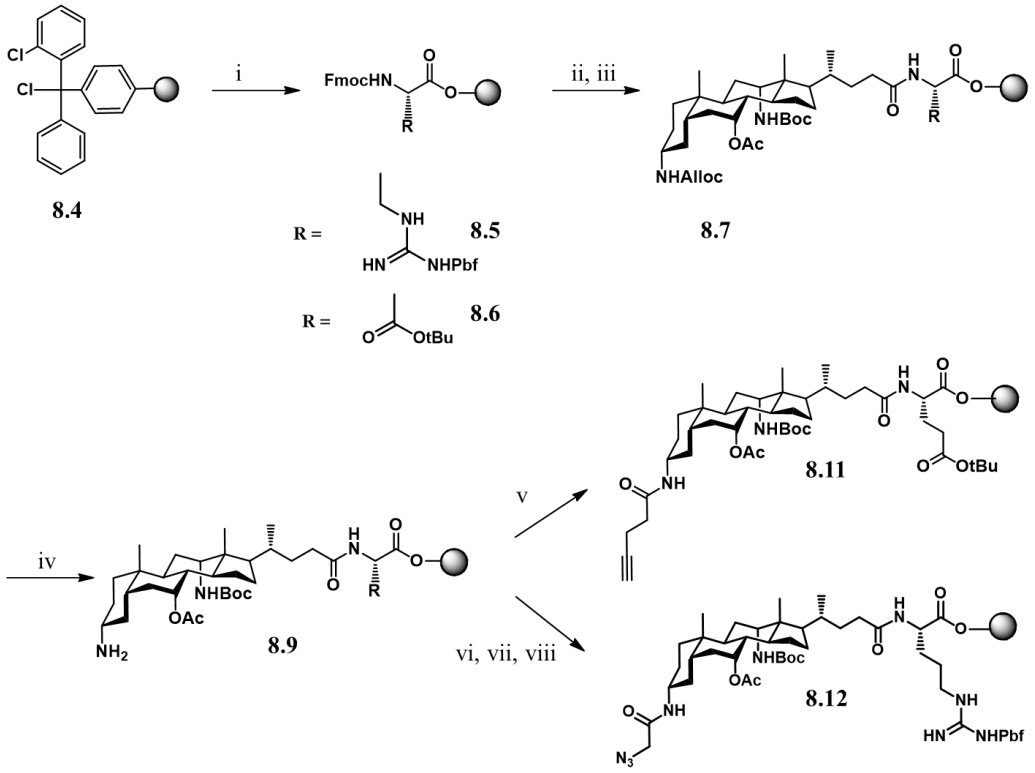


Figure 8.5. Synthetic scheme for the obtention of the alkyne and azide parts **8.11** and **8.12** respectively. i) Fmoc-aa-OH, DIPEA, DCM (dry), RT 3h. ii) 20% piperidine/DMF, RT 2 x 20 min. iii) scaffold, PyBOP, DIPEA, NMR, RT 24h. iv) $PhSiH_3$, $Pd(PPh_3)_4$, DCM, RT 3 x 1h. v) 4-pentynoic acid, PyBOP, DIPEA, NMR RT 3h. vi) Fmoc-Gly-OH, PyBOP, DIPEA, NMP RT 3h. vii) 20% piperidine/NMP 2 x 20 min. viii) TfN_3, $NaHCO_3$, $CuSO_4 \cdot 5H_2O$, MeOH, RT 24h.

Cleavage of both counterparts and conjugation via CuAAC afforded the cage-like receptor. Final deprotection of the Boc, tBu and Pbf groups and lyophilization with $NaHCO_3/Na_2CO_3$ allowed formation of the correctly charged product **8.13.** This charged state could enable the formation of a salt bridge between the side chains of arginine and glutamine that warranted the closure of the structure in a cage conformation that could act as the EDC's receptor (figure 8.6).

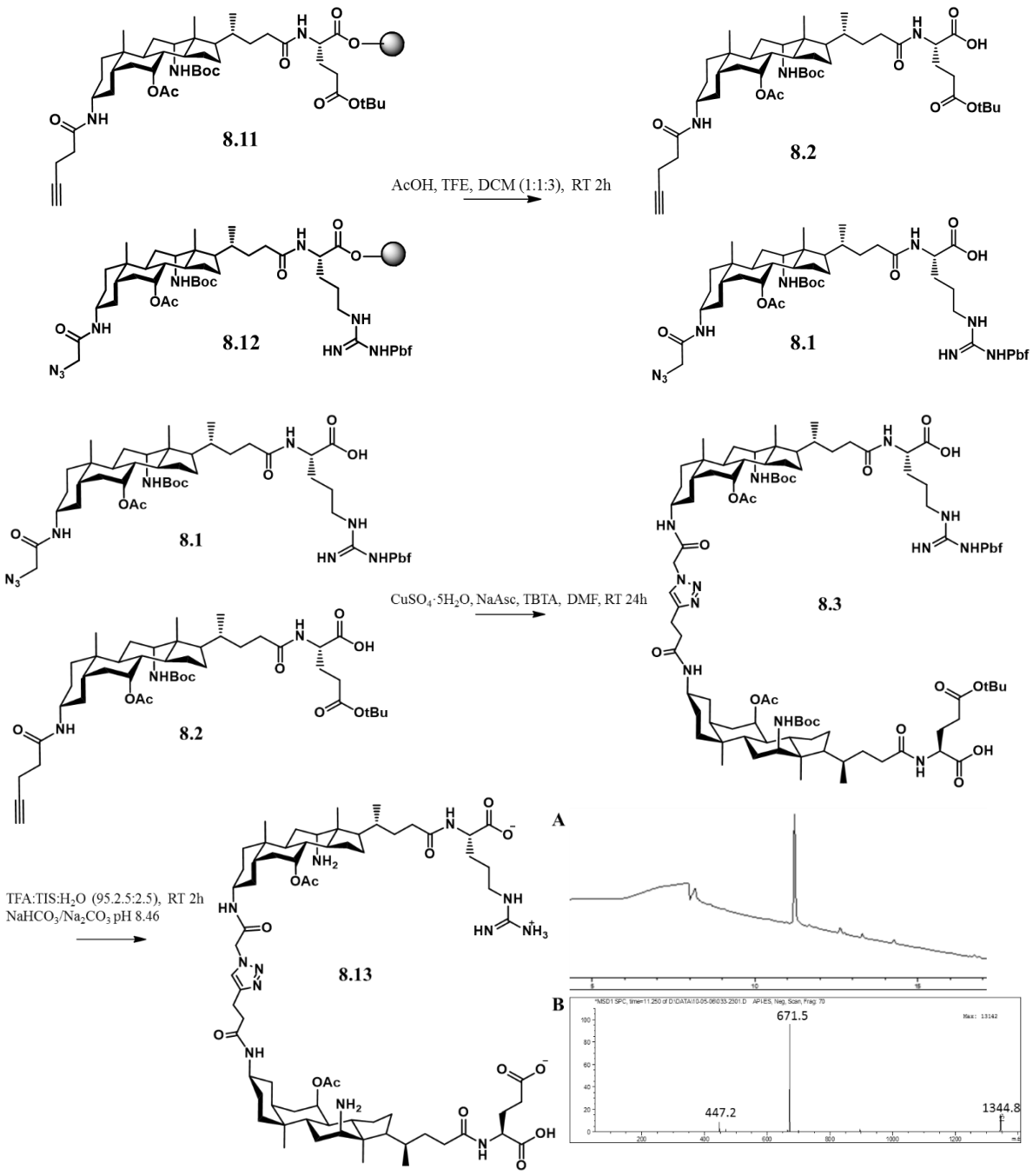


Figure 8.6. Cleavage, conjugation and deprotection towards the obtention of compound **8.13**. A) LC-MS chromatogram of pure compound **8.13** before lyophilization (Luna C18 (250x4.6 mm, 5 µm at 35°C; 0-100% ACN in 15 min). B) Corresponding ESI-MS spectrum at 11.250 min. M = 1345.7. Found = 1344.8 [M-H$^-$] 671.5 [M/2 – H$^-$] and 447.2 [M/3 – H$^-$].

8.1.2. Conclusion

The synthesis of a cage-like receptor for the accommodation of an estradiol molecule has been optimized. Unfortunately, NMR studies showed that there was no affinity between both counterparts. For further optimization of the design, we propose the incorporation of a spacer between arginine and glutamine and the scaffold, in order to enlarge the cavity for the proper interaction with 17β-estradiol.

8.2 Development of water-soluble NHC-Cu Catalysts: Applications in click chemistry, bioconjugation and mechanistic analysis.

This section has been published in Organic and Biomolecular Chemistry in 2014 (Online manuscript, DOI: 10.1039/C4OB01350F) and as authors Heriberto Díaz Velázquez, Yara Ruiz García, Matthias Vandichel, Annemieke Madder and Francis Verpoort [247].

This project was developed in collaboration with Prof. Francis Verpoort and his PhD student Heriberto Diaz to develop water-soluble NHC-Cu Catalysts: Applications in Click Chemistry, Bioconjugation and Mechanistic Analysis. The next paragraph serves as an introduction to this section, adapted from the article we published[248]. Although several catalysts have been developed for the Copper catalyzed Alkyne-Azide Cycloaddition reactions (CuAAC, better known as the Click reaction), in most cases they require the presence of co-catalysts such as auxiliary ligands, bases – mainly amines and reducing or oxidizing agents depending on the Cu source used, in order to enhance their catalytic activity, although in recent years single Cu salts have also been succesful[249,250]. Generally high Cu loading must be applied to ensure good performance of the catalyst[251]. Several Cu(I) complexes holding *N*-heterocyclic carbene (NHC) ligands are reported as catalysts for the Huisgen Cycloaddition reaction, however the catalysis is carried out at elevated temperatures and in the presence of organic solvents, under two-phase systems (when both reactants are not soluble) and only the 1,4-disubstituted triazoles are generated[252,253]. Moreover, their activity in solution-phase is significantly lower than the other catalytic systems[5]. Here we report an unique Cu(I) based catalyst that enables the synthesis of triazoles with different substitution patterns, *e.g.* 1-, 4-, 1,4- or 1,4,5-substituted triazoles and at the same time, able to generate triazoles from acetylene in water, able to produce tosyl acetamides in aqueous media (figure 8.7).

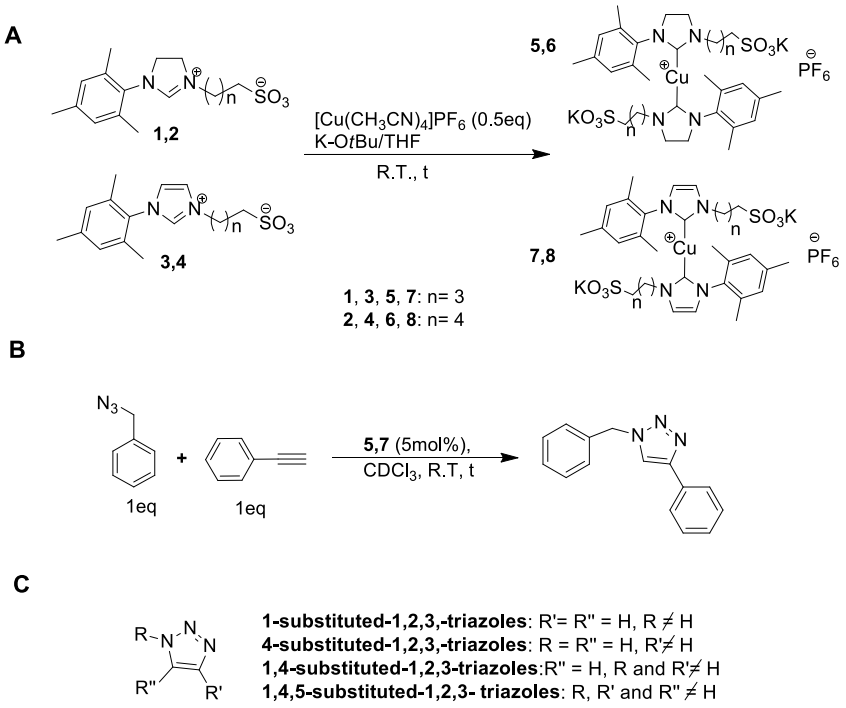


Figure 8.7. A Synthesis of catalysts 5-8; B. Standard Click reaction monitored by ^{1}H-NMR using catalysts 5 and 7; C Triazoles synthesized in this work.

In the last decade, bioconjugation has become a common tool for chemists and biochemists as a result of the increase in peptide and protein research and development. This covalent functionalization of these biomacromolecules under physiological conditions gives rise to new areas of research such as drug discovery, high-throughput screening and *in vivo* testing. Therefore, many ligation methodologies have been studied and optimized[254,255]. The development of Click chemistry has caused an emerged interest in the functionalization of peptides with alkynes and azides, resulting in a broad variety of building blocks such as unnatural amino acids being currently commercially available. In practice, the application of CuAAC in peptide chemistry requires extra efforts for the optimization of the reaction conditions[255]. Traces of O_2 can be reduced by Cu^{I} and sodium ascorbate into H_2O_2, a reactive oxygen species that can induce degradation of amino acids and cleavage of the polypeptide chain[127,256–258]. Thus, a catalytic system able to perform click reactions in biomacromolecules under reductant-free aqueous conditions will lead to a major step forward in bioconjugation chemistry. Furthermore, bioconjugation impose specific requests, such as exquisite chemoselectivity, biocompatibility, and the ability to work at low temperature still enabling fast transformations.

To date the CuAAC reaction in the absence of an accelerating ligand is simply too slow[259]. In this context, we have studied the performance of our water-soluble NHC-Cu (I) catalysts for CuAAC reactions with unprotected peptides in aqueous conditions. We have chosen the DNA binding domain of the transcription factor protein GCN4, belonging to the bZIP Leucine Zipper family as a substrate for the reaction[53]. Indeed, modification of the GCN4 binding domain in the design of transcription factor models has been of particular

interest for gene therapy[260]. Surprisingly, there are no studies so far on the application of click chemistry on the modified GCN4 binding domain for conjugation purposes. Here, we report the successful functionalization of the peptide comprising the basic region DNA binding domain with an organic azide via click chemistry using the sulfonated NHC-Cu (I) catalysts.

To further expand and illustrate the scope and potential of the developed catalysts, bioconjugation experiments with alkyne and/or azide tagged peptides were carried out. Hereto, two peptides, one 10 mer and one 23 mer, comprising part of the N-terminal basic region of the Transcription Factor protein GCN4, were synthesized on 2-chlorotrityl chloride polystyrene resin. Propargyl glycine was incorporated as the first, C-terminal amino acid residue (figure 8.8) to provide for the alkyne functionality for further click reaction with benzylazide. Click reaction was carried out at ambient temperature in a deoxygenated mixture of water and hexafluoroisopropanol to disrupt H-bonding interactions within the peptide using 1.5 equivalents of catalyst (6 for modification of 10mer GCN4 and 7 for 23mer GCN4 due to its higher stability) under anaerobic conditions overnight. The resulting modified peptides were purified by High Performance Liquid Chromatography (HPLC) and further characterized by Electrospray Ionization Liquid Chromatography connected to a Mass Spectrometer (LC-MS-ESI) and MALDI-TOF. Generation of the Cu-acetylide intermediate in such an environment is highly challenging, therefore a slight excess of catalyst 6 is needed. In current context, it is further worth noticing that both sequences are rich in arginine residues, of which the guanidinium side chain can easily coordinate with copper and inhibit the reaction.

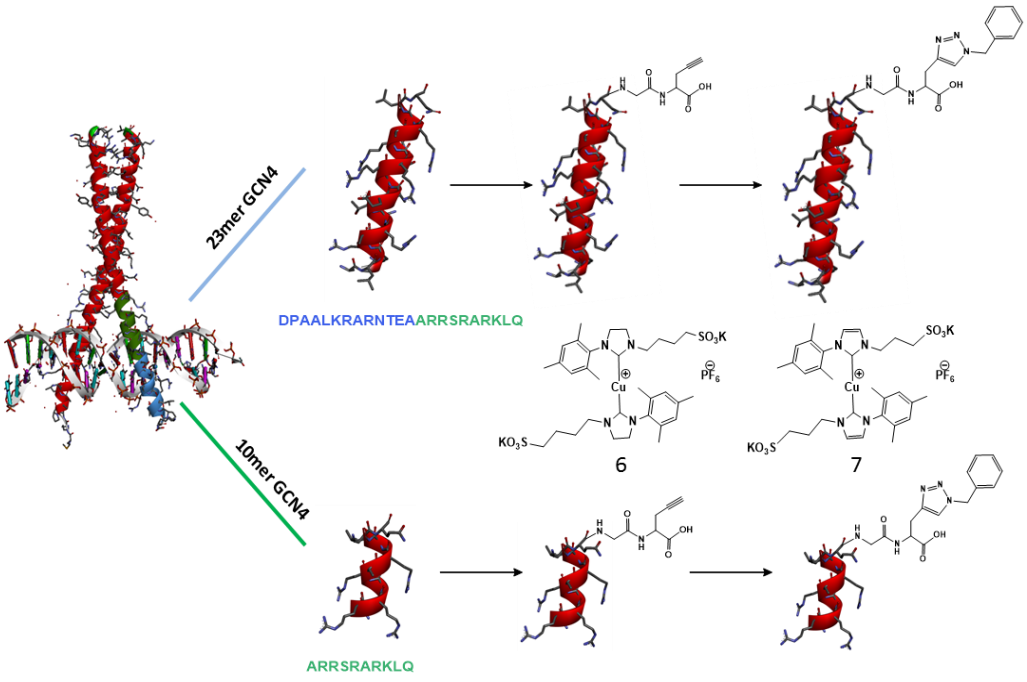


Figure 8.8. The transcription factor protein GCN4, belonging to the bZIP Leucine Zipper family and the derived model peptide comprising the DNA binding domain.

Functionalization of a peptide with an alkyne allows cycloaddition-based modification with a non-peptide azide moiety. These experiments underscore the potential for application of our water-soluble $(NHC)_2Cu$

(I) catalysts 6 and 7 in the functionalization of deprotected peptides applying the CuAAC strategy. We haver further shown the use of sulfonate functionalized bis-NHC-Cu(I) water-soluble complexes in bioconjugation experiments, attaining for the first time the functionalization of an unprotected 25-aminoacid chain peptide using low catalyst loading, which otherwise would need a much higher amount of Cu(I) to achieve high efficiency due to the chelating properties of certain aminoacid residues.

CHAPTER 9

CONCLUSIONS & PERSPECTIVES

Through this work, we have achieved the successful synthesis of major groove DNA binders based on bZIP and bHLH transcription factors for the sequence-selective recognition of DNA. Different models have been designed and synthesized and have given us insight in the structural features that influence the DNA binding capabilities of artificial binders. Moreover, we have also studied the cellular uptake of these constructs.

In parallel, we have also achieved the development of peptidosteroid conjugates by means of a new peptide chemical ligation strategy. Finally, we have also taken the first steps towards the development of synthetic receptors of phosphatidylserine and its application on the early detection of apoptosis and as flippase mimics.

9.1 Detection of phosphatidylserine

Receptors with different cavities to accommodate phosphate anions were synthesized and tested for the detection of PS. The acyclic receptor with aliphatic chains (**1.6**) proved to be the best binder, probably due to the aliphatic chains that contribute via hydrophobic interactions with the lipid tails of PS.

Prior to analysis in cells, the receptor should be tested for cell penetration. Detection of PS on the outer part of the membrane requires a compound that will not be able to transport the phospholipid across the cellular membrane. Although this could be possible, this negative result can be transferred to the development of synthetic flippases, the enzymes that transport PS from the outer face of the membrane to the inner face.

In order to assess the externalization of PS in the cell, we envisage that a fluorophore will be needed for further analysis by flow cytometry. The special features of the steroid scaffold allow the assembling of this moiety through the C24 terminus (figure 9.1).

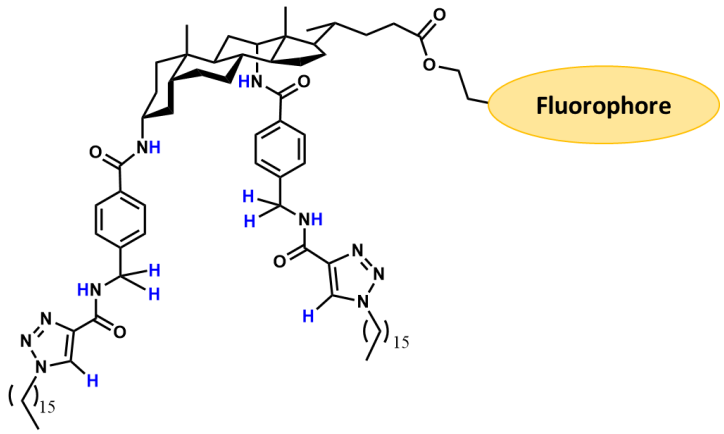


Figure 9.1. Structure of a fluorescently labelled receptor.

9.2 Transcription factor models and cell uptake

From the four peptidosteroid models of the GCN4 bZIP transcription factor presented in chapter 2, the one with a longer and more flexible linker proved to be the best synthetic DNA binder. This could be attributed to the fact that the longer linker allowed the construct to grip the major groove of the DNA like a pair of tweezers which was not possible in the case of the less flexible linkers. Its binding affinity was comparable with the earlier described models of GCN4 TFs, however, the synthetic route towards obtaining the current construct was considerably simplified. We also illustrated an unprecedented, optimized CuAAC strategy to conjugate relatively long, deprotected peptides to bile acid scaffolds in a convergent manner. In addition, our best DNA binding construct showed a clearly enhanced cell uptake when tested against RAW264.7 mouse macrophages illustrating the particular properties of the cholic acid scaffold as a versatile dimerizing unit with low cytotoxicity at concentrations which allowed sequence specific DNA binding (figure 9.2).

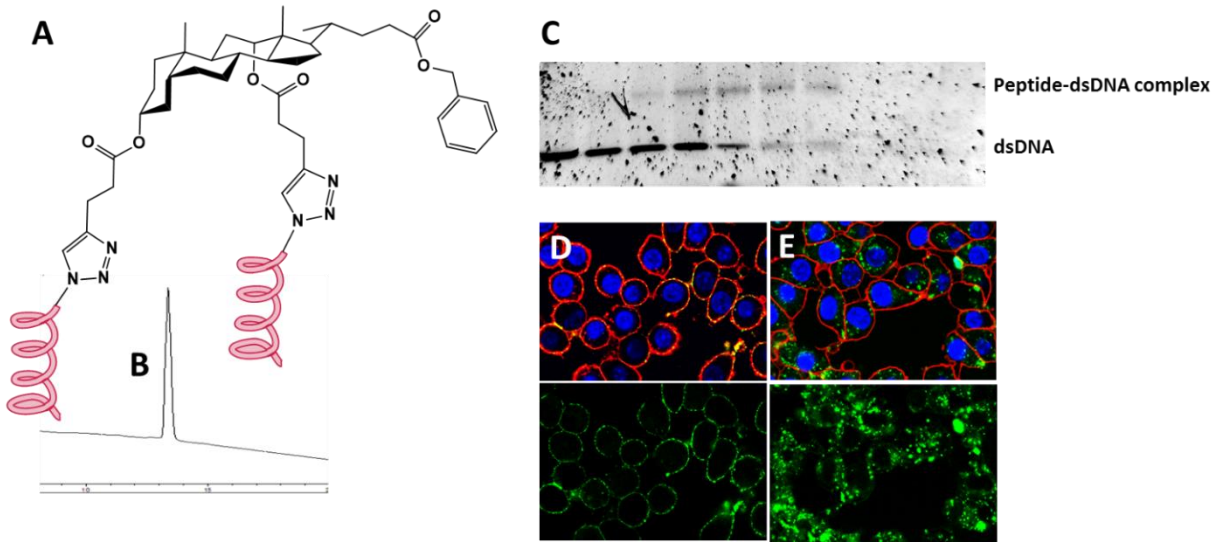


Figure 9.2: Characterization of (A) final synthetic construct using (B) HPLC (C) DNA binding (D) Confocal microscopy at 4°C and (E) 37°C.

In order to obtain further insight in the influence of the structural features for DNA recognition in synthetic binders we explored the potential of peptide-cyclodextrin conjugates for sequence-selective DNA recognition. This was achieved by the use of α, β and γ-cyclodextrin derivatives as scaffolds for the appendage of the peptides by CuAAC. Though examples exist of CuAAC reactions with CD derivatives, to the best of our knowledge long, deprotected peptides have so far not been conjugated to cyclodextrins via CuAAC. We here presented optimized conditions for the anchoring of such long peptides to cyclodextrin units via click chemistry. Our results indicated the usefulness of an optimized dimerization configuration between both peptides in artificial TF models. Indeed it was shown that the distance between the anchoring points influenced DNA binding (figure 9.3).

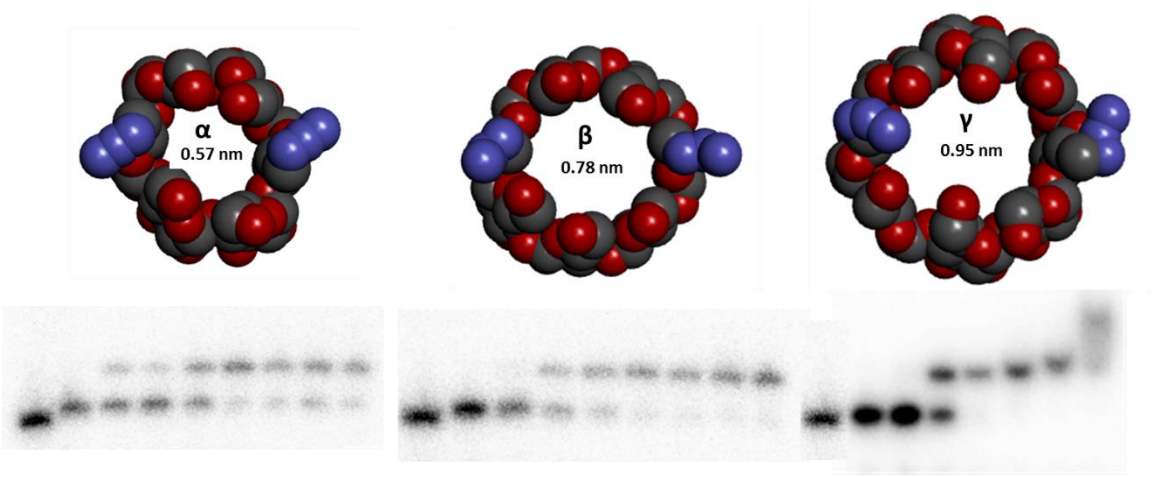


Figure 9.3: EMSA for α, β and γ-cyclodextrin-peptide conjugates.

To conclude, the potential of a reduced-size bHLH-ZIP cMyc/Max TF to recognize its dsDNA binding sequence was tested. The results showed that an optimized and miniaturized version of the protein contained the necessary features to form a complex with dsDNA in a sequence-specific manner (figure 9.4).

In addition, we showed the possibility of selective functionalization of steroid-based scaffolds to synthesize heterodimeric peptide-based conjugates, broadening the application of the current methodology to other families of TFs, such as bZIP Fos/Jun TF. In addition to the above, our design overcame the difficulties of the decoration of building blocks by the well-known SPPS, since it is based solely on conjugation strategies in solution phase. This can be considered the first model of this oncoprotein accessible by synthetic methods.

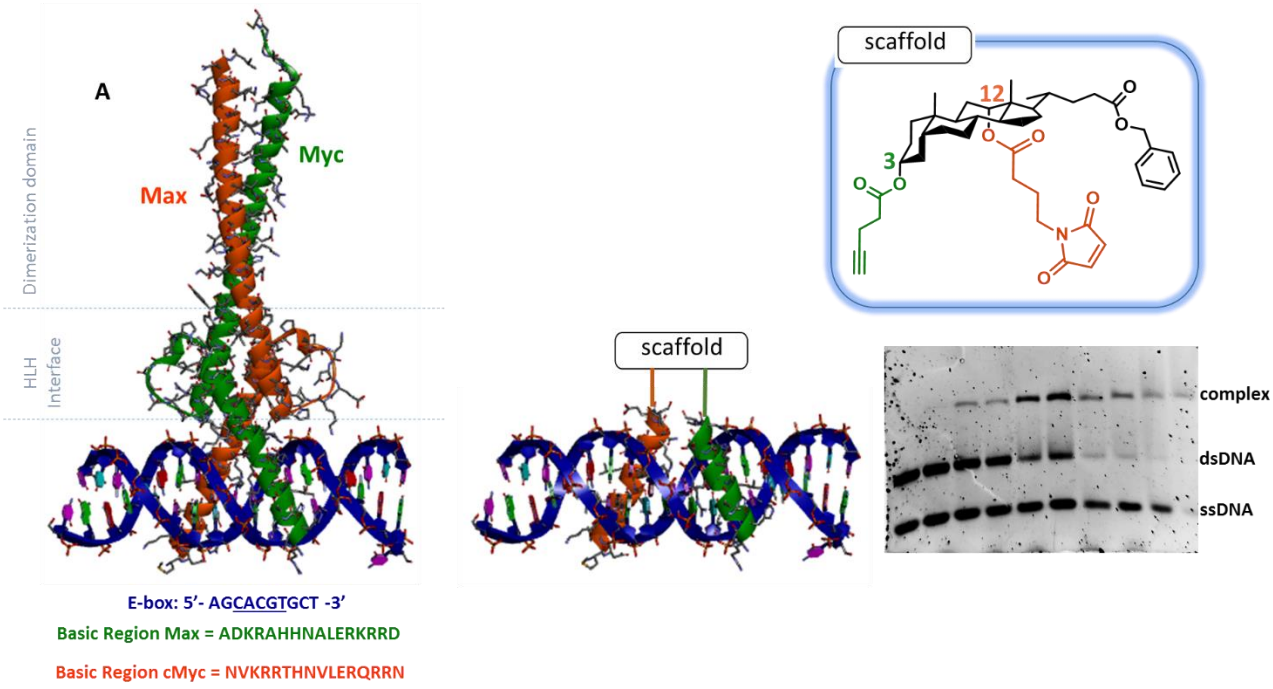


Figure 9.4: Crystal structure of the cMyc-Max protein bound to DNA, designed synthetic mimic and DNA binding pattern.

9.3 Resin to Resin Transfer Reaction (RRTR)

In chapter 6 we described the last steps and optimisation of a new peptide chemical ligation strategy based on the resin-to-resin conjugation of the peptide and a steroid scaffold (figure 9.5). Our strategy can be useful for solid-phase conjugations amenable to library formats and applicable in areas as diverse as protein models, receptor enzyme design, chemical biology and classical medicinal chemistry. It can also be compatible with synthesis under microwave conditions.

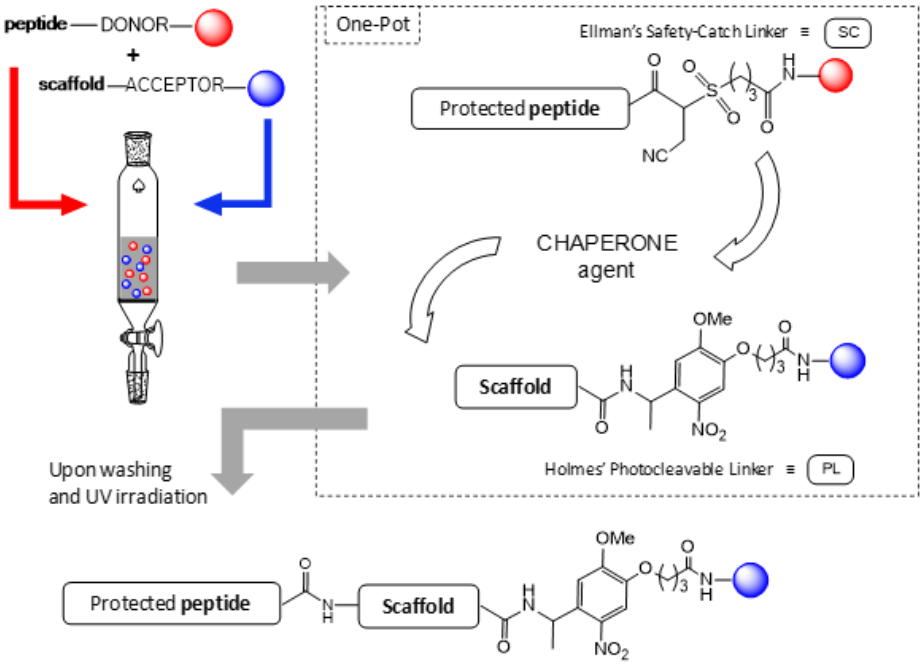


Figure 9.5: Overview of the RRTR.

9.4 Perspectives

In the current thesis, the DNA binding mimics of the basic region of the GCN4 Leucine zipper TF were the key compounds that were tested for possible future applications for DNA binding in cells. Despite the high binding affinity of the mimics towards dsDNA as shown via EMSA, there is still work to be done if this system has to succeed in vivo. Since there was no uptake in the nucleus for any of the constructs, we believe the first next step should be adding a short NLS to the construct which may enhance uptake in the nucleus as is observed for the Tat peptide[261]. For this purpose the –COOH functionality of the scaffold is the synthetically most accessible point.

In chapter 7, we described our perspectives for the employment of synthetic DNA binders as artificial nucleases by the incorporation of cleaving moieties on the side chains of the peptide. We strongly believe that joint efforts between experts on peptide design for DNA recognition and phosphate hydrolysis will result in constructs with the potential to cleave sequence-selectively dsDNA on the major groove via hydrolysis of the phosphate group in the DNA backbone. Due to the on-going research on development of Zn(II) complexes for phosphate cleavage and mechanistic studies, different features towards more efficient catalysts have been revealed. Therefore, future attempts should provide a suitable catalyst for the challenging cleavage of a phosphodiester bond within dsDNA.

PART II:

EXPERIMENTAL SECTION

CHAPTER 10:

GENERAL MATERIALS AND METHODS

10.1 Products

10.1.1 Solvents

All organic solvents were purchased from commercial suppliers and used without further purification or drying. DMF extra dry (with molecular sieves, water < 50 ppm) and other dry solvents with sieves namely tetrahydrofuran (THF), pyridine and methanol (MeOH) were acquired from ACROS Organics. Anhydrous N-Methyl-2-pyrrolidone (NMP) was acquired from Fluka. N,N-Dimethylformamide (DMF) an NMP (peptide synthesis grade) were purchased from Biosolve. Ethyl Acetate, Acetonitrile, Methanol, Diethyl Ether, Et_3N and DIPEA supplied as extra dry, redistilled, were purchased from Sigma Aldrich. HPLC grade quality Hexane and Chloroform were purchased from Fisher Scientific. Deuterated solvents $CDCl_3$ (99.8% atom D), CD_2Cl_2 (99.9% atom D) and CD_3OD (99.8% atom D) were obtained from Sigma-Aldrich. Water with the Milli-Q grade standard was obtained in-house either from a Millipore ROs 5 purification system or a Sartorius Arium 611 DI.

TentaGel NH_2 (0.28 mmol/g, 90 μm) was obtained from Merck Novabiochem and IRIS Biotech GmbH. preloaded H-Gly-Sulfamylbutyryl NovaSyn TG resin (0.24 mmol/g, 90 μm) were obtained from Merck Novabiochem.

10.1.2 Reagents for peptide synthesis

Rink-Amide ChemMatrix (100-200 μm, manufacturer's loading: 0.52-0.54 mmol/g) was obtained from Biotage. 2-Chlorotrityl Chloride resin (100-200 manufacturer's loading 1.6 mmol/g) was purchased from Iris Biotech GmbH.

Fmoc-Propargylglycine-OH, Fmoc-Azidohomoalanine-OH, PyBOP and HBTU coupling reagents were obtained from either Merck Novabiochem or IRIS Biotech GmbH, while HATU (purum ≥ 98.0 %) was acquired from Fluka. TFA was purchased from Iris Biotech GmbH. NHS-Fluorescein was purchased from Thermo Scientific. Triisopropylsilane (TIS) and tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were purchased from Sigma Aldrich and Phenylsilane was purchased from ACROS Organics.The Nα-Fmoc protected amino acids were purchased at Merck Novabiochem, IRIS Biotech GmbH and Fluka, or supplied by MultiSynTech GmbH. All chiral α-amino acids used in this paper possessed the L configuration. Throughout this work, residues with standard acid-sensitive side-chain PGs were used: Cys(Trt) [C], Asp(OtBu) [D], Glu(OtBu) [E], His(Trt) [H], Lys(Boc) [K], Asn(Trt) [N], Gln(Trt) [Q], Arg(Pbf) [R], Ser(tBu) [S], Thr(tBu) [T].

All reagents were acquired from commercial sources and used without prior purification.

10.1.4 Reagents for EMSA

Tris-borate-EDTA buffer 10x pH 8.3 (TBE buffer), ammonium persulfate (APS), tetramethylethylenediamine (TEMED), were obtained from Merck. SYBR Gold Nucleic acid gel stain was purchased from Life Technologies.

10.1.5 Oligonucleotides

CRE	5’ – CGG ATG ACG TCA TTT TTT TTC – 3’
CRE complement	5’ – GAA AAA AAA TGA CGT CAT CCG – 3’
Random	5’ – GCG CGA GAA GGA AAG AAA GCC GG – 3’
Random Complement	5’ – CCG GCT TTC TTT CCT TCT CGC GC – 3’
E-Box	5’ – CTA CTA GCA CGT GCT AGT AG – 3’
E-Box Complement	5’ – CTA CTA GCA CGT GCT AGT AG – 3’

All oligonucleotides used were commercially purchased from Eurogentec (HPLC purified using RP-cartridge-Gold, 200 nm scale) and were used as such.

10.1.6 Other reagents, chemicals and materials

1.11-diazido-3,6,9, trioxaundecane was purchased from Toronto Research Chemicals. Anilinium *p*-toluenesulfinate, which was synthesized according to Nagakura et al.[1] HOBt (98 %) and HOAt (98 %) and benzylamine ≥ 99.5 % pure were purchased from ACROS Organics. NaSPh was technical grade (90 %), purchased from Sigma-Aldrich. HOBt, HOAt and NaSPh were dried under high vacuum prior to use. Tetrakis-(triphenylphosphin)-palladium(0), 99 % pure was purchased from Sigma. Tributyltin hydride from Fluka was purum ≥ 96.0 %. Morpholine from Sigma-Aldrich was ≥ 99 %.

Merck silica refers to: Kieselgel Merck Typ 9385 230-400 mesh, 60 Angström, and specifications of Biosolve silica are: 60 Å, 0.063-0.200 mm.

Occasional sonication to aid dissolution/suspension/homogeneity was performed in a Branson 2510 water bath. Falcon tubes were composed of high-clarity polypropylene, with a capacity of either 15 mL, (17 x 120 mm; BD Falcon), or 50 mL (VWR).

10.2 Automated Peptide Synthesis

Automated peptide syntheses were performed on a fully-automated SYRO Multiple Peptide Synthesizer robot, equipped with a vortexing unit for the 24-reactor block (MultiSynTech GmbH). Reactions were open to the atmosphere and were executed at ambient temperature with protection from light.

Standard Fmoc/tBu chemistry was used for SPPS with HBTU as a coupling reagent and 20-40% piperidine in NMP for Fmoc deprotection. The resin is preswollen in NMP for 10 min and then filtered off. The following protocols were used for Fmoc deprotection and coupling:

[1] Honda, M.; Morita, H.; Nagakura, I. *J. Org. Chem.* **1997**, *62*, 8932-8936. Deprotection of allyl groups with sulfinic acids and palladium catalyst.

Fmoc deprotection: Fmoc deprotection is performed by adding a solution of 20-40% piperidine in NMP to the resin and shaking for 2 min, 5 min & 15 min durations. After each addition and shaking cycle, the resin is filtered off and washed with NMP (6 x 45 s).

Coupling: Single couplings using 10 equivalents of a 0.5 M solution of the Nα protected amino acid in NMP/DMF, 10 equivalents of a 0.5 M solution of HBTU in DMF and 10 equivalents of a 2.0 M solution of DIPEA in NMP which are added to the resin in that order. The reaction mixture is shaken for 1 hour. The resin is filtered off and washed with NMP (9 x 2 min).

Capping: 4-Acetamidobenzoic acid used for capping was coupled using the protocol in *coupling* above. In case of fluorescein a manual coupling was performed as described in *coupling* above.

10.3 Analysis & Equipment

10.3.1 HPLC analysis and purification

Reversed-Phase HPLC analysis and purification was performed on an Agilent 1100 Series instrument with diode array detector (at 214, 254, 280, 310, 360 nm), equipped with a Phenomenex Luna C18 100 Å column (250 x 4.6 mm, 5 µm, at 35 °C) or a Phenomenex Jupiter C4 300 Å column (250 x 4.6 mm, 5 µm, at 35 °C). Linear gradient elutions were performed by flushing 2 min with A followed by 0 to 100% buffer B in 15 minutes and finally by a 5 min flushing with B using a binary solvent system composed of buffer A: 0.1% TFA in H_2O and B: MeCN with a flow of 1.0 mL/min at 35°C.

Semi-prep purification was performed using a Phenomenex Luna C18 100 Å column (250 x 10 mm, 5 µm, at 35 °C) with a flow rate of 4.5 mL/min and prep purification was performed using a Phenomenex Luna C18 100 Å column (250 x 21 mm, 5 µm, at 35 °C) with a flow rate of 17.5 mL/min. Gradients were varied depending on the separation.

10.3.2 LC-MS analysis

LC-TIC-MS data (reversed phase) were recorded on an Agilent 1100 Series instrument with diode array detector (at 214, 254, 280, 310, 360 nm), equipped with a Phenomenex Kinetex C18 100 Å (150 x 4.6 mm, 5 µm, at 35 °C), hyphenated to an Agilent ESI-single quadrupole MS detector type VL. Mass detection operated in the positive mode. Linear gradient elutions were performed by flushing 0.5 min with A followed by 0 to 100% buffer B in 6 minutes and finally by a 2 min flushing with B using a binary solvent system composed of buffer A: 0.1% HCOOH in H_2O and B: MeCN with a flow of 1.0 mL/min at 35°C. Pre and post flushing is included in the LC-MS and RP-HPLC results.

10.3.3 ESI-MS analysis

ESI-MS spectra were recorded on a quadrupole ion trap LC mass spectrometer (Thermo Finnigan MAT LCQ), equipped with electrospray ionization. MeOH/H_2O (4/1 ± 0.1 % formic acid) was used as carrier solution. All reported data were collected in the positive mode, at 250 °C.

10.3.4 MALDI-TOF

MALDI-TOF-MS data were acquired on an Applied Biosystems Voyager-DE STR Biospectrometry Workstation, equipped with a high performance nitrogen laser (337 nm). All spectra were recorded in the positive and reflector mode, with delayed extraction. A solution of 4-5 mg α-cyano-4-hydroxycinnamic acid in 500µl MeCN, 490µl mQ, 10µl 1M ammoniumcitrate, 1µl TFA was used as matrix for MALDI-TOF-MS.

10.3.5 NMR Spectra

NMR spectra were recorded at room temperature on a Bruker Avance-300 spectrometer at 300 MHz for ^{1}H NMR and on a Bruker Avance-500 spectrometer at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C spectra. Chemical shifts (δ units) are expressed in parts per million (ppm) relative to tetramethylsilane (TMS) and the internal solvent peak was used for calibration. Wherever peak multiplicities have been reported, the following abbreviations have been used: s, singlet; d, doublet; t, triplet; m, multiplet; br, broad. Coupling constants (J values) are expressed in Hertz (Hz). The Attached Proton Test (APT) technique was used to assign ^{13}C peaks (C, CH, CH_2, CH_3). All ^{1}H spectra of deoxycholic acid derived products contain a region of high signal overlap between 1.00 and 2.00 ppm. Therefore, the detailed analysis of this envelop is almost impossible and often not mentioned or reported as a series of multiplets. Fortunately, in the analysis of the synthesized scaffolds all necessary information was shifted downfield from this region.

Concentration of final products were determined by ^{1}H-NMR ERETIC (Electronic REference To access In vivo Concentrations)[91] with a high-resolution 500 MHz NMR-spectrometer.

10.4 Electrophoretic Mobility Assay (EMSA)

10.4.1 Sample preparation

The following stock solutions were prepared (fresh each time, except for DNA and peptide):

DNA: A concentration of 1.67 µM prepared from the required DNA and its complement. DNA solutions were made (commercially purchased) by diluting with 20 µL 0.5 M Tris, pH = 8, 40 µL 2.5 M NaCl, 40 µL 0.025 M EDTA and then adding milliQ water such that the total volume is 1 mL. The DNA was annealed by heating in a Thermomixer from room temperature to 95°C and maintaining the temperature at 95°C in total time of 24 min. The machine was then turned off and the sample was allowed to cool down slowly.

In order to ensure physiological conditions, proper solubility and optimal charges in both peptide and DNA, a loading buffer was prepared, consisting in 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M and 40 µL EDTA 0.025 M. The main purpose of using loading buffers is to ensure good solubility and ionization as is required for gel electrophoresis. A high amount of salt concentration can however significantly decrease DNA binding due to an increase in ionic strength and subsequent decrease in the peptide-DNA interactions. EDTA is used to bind to stray metal ions such as Mg^{2+}, Ca^{2+} but does not

significantly affect peptide-DNA interactions. Sucrose ensures that the loading solution is heavier than the buffer used to run the gels so that the solution does not flow over the well.

Sucrose: 30% sucrose in mQ (300 mg/mL).

Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water.

Loading mixture: The loading mixture comprised of: 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide. The loading mixture was prepared only 1-2 hr prior to running of gels and kept on ice as soon as ready.

10.4.2 Preparation of Gels (for 2 Gels)

In a clean falcon tube the following were added (in given order): 15.595 mL mQ, 0.4 mL TBE, 4.005 mL of 40% acrylamide solution, 200 µL APS (10% w/w in mQ). The solution was mixed by sonication to remove any air bubbles and cooled to 0°C (1 h under ice). 20 µL of TEMED was then added to the mixture and was again mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (~1 h).

10.4.3 Gel Electrophoresis

A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2x TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. Instrument settings: 150 V, 100 mA, 19 W for 30 mins at 4°C. The wells were washed after the pre-run. 5 µL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 mins at 4°C.

10.4.4 Staining of gels

After the run, the gels were removed from the glass and were stained using 100 mL of 0.2x TBE buffer + 10 µL Sybr Gold (commercially purchased stock solution 10,000X in DMSO). The gels were then washed twice with mQ and gently placed under a UV lamp (dark room) to observe the gel pattern.

10.5 Electrophoretic Mobility Shift Assay (EMSA) for quantification of the dissociation constant

The EMSA for the quantification of the dissociation constant were carried out by Vladimir Pabon at Karolinska Institute in Sweden.

10.5.1 Preparation of ^{32}P-labeled dsDNA target

The required oligonucleotide was 5'-labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase (Fermentas) according to the manufacturer's protocol and then purified using QIAquick Nucleotide Removal Kit (Qiagen). The 5'-end labeled pyrimidine oligonucleotide was annealed with the unlabeled complementary

strand. An amount of 5nM dsDNA was prepared diluting 20 µL 0.5 M Tris, pH = 8, 40 µL 2.5 M NaCl, 40 µL 0.025 M EDTA and then adding milliQ water such that the total volume is 1 mL. The DNA was annealed in a heat block by heating from 95°C during 5 minutes and then slow cooling to room temperature.

Loading buffer: 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M.

Sucrose: 30% sucrose in mQ (300 mg/mL)

Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water.

Loading mixture: The loading mixture comprised of: 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide. The loading mixture was prepared only 1 hour prior to running of gels and kept on ice as soon as ready.

10.5.2 Preparation of Gels (for 1 Gel)

In a clean glass beaker the following were added (in given order): 21.57 mL mQ, 0.6 mL TBE, 7.5 mL of 40% acrylamide solution (29:1), 0.3 mL APS (10% w/w in mQ) and 30 µL of TEMED was then added to the mixture and mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (40 minutes).

10.5.3 Electrophoresis

A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2x TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. The wells were washed after the pre-run. Instrument settings: 150 V, 100 mA, 19 W for 30 minutes with circulation water-cooling. 5 µL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 minutes with circulation water-cooling.

The gels were frozen and analyzed by phosphor imaging using Molecular Imager FX and the data were processed using Quantity One software (BioRad).

CHAPTER 11:

CHAPTER-WISE EXPERIMENTAL SECTION

Chapter 11.1

11.1.1 Synthesis and characterization of compounds

Synthesis of Hexadecylazide[262]

11.1.1 **11.1.2**

Hexadecylamine **11.1.1** (100 mg, 0.41 mmol) was dissolved in MeOH (3 mL) and treated with K_2CO_3 (114.48 mg, 0.828 mmol) followed by $CuSO_4 \cdot 5H_2O$ (1 mg, 0.0041 mmol). To this mixture, imidazole-1-sulfonyl azide·HCl (105.04 mg, 0.501 mmol) dissolved in MeOH (2 mL) was added and the reaction was stirred at room temperature under argon for 2h. The reaction was quenched by the addition of H_2O (3 mL). This was acidified with AcOH and extracted with hexane (2x4 mL). The combined organic layers were dried under $MgSO_4$ and evaporated to give a colorless glue **11.1.2** (105 mg, 95%).

^{1}H-NMR (300 MHz, $CDCl_3$); δ 3.25 (m, 2 H), 1.6 (m, 2 H), 1.25 (br m, 26 H), 0.85 (t, 2 H). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 50.4854 (CH_2), 30.9200 (CH_2), 29.3015 (CH_2), 28.6783 (CH_2), 28.6519 (CH_2), 28.6177 (CH_2), 28.5380 (CH_2), 28.4780 (CH_2), 28.3559 (CH_2), 28.2332 (CH_2), 28.1522 (CH_2), 27.8352 (CH_2), 25.7130 (CH_2), 21.6831 (CH_2), 21.0841 (CH_2), 13.1012 (CH_3).

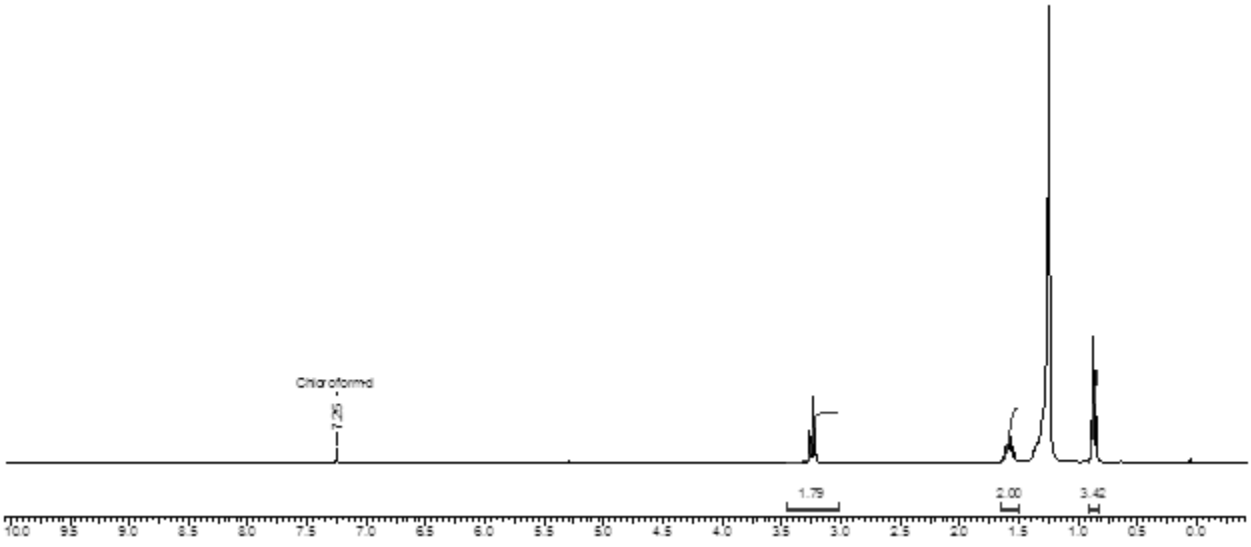


Figure 11.1.1. ^{1}H NMR spectrum of compound **11.1.2**.

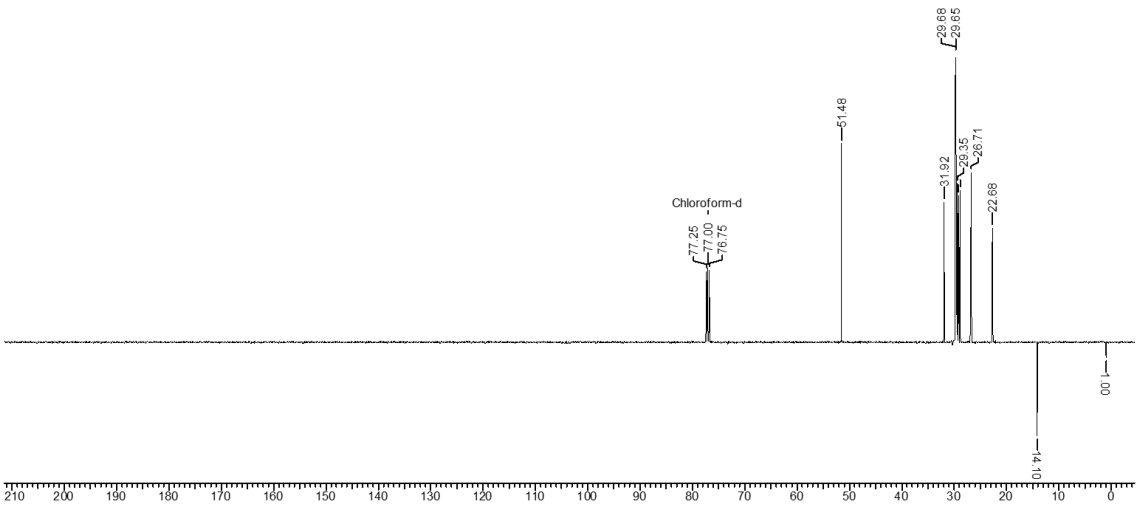
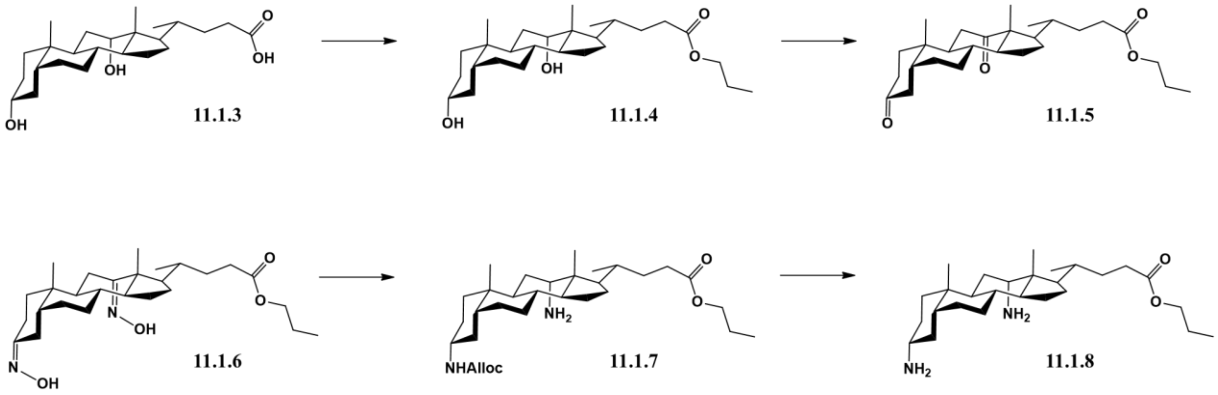


Figure 11.1.2. ^{13}C NMR spectrum of compound **11.1.2**.

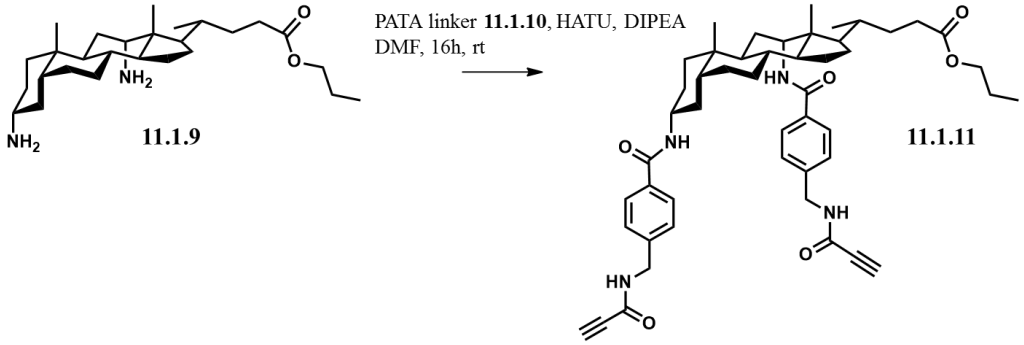
** The Diamino scaffold **11.1.8** was synthesized according to reference [48,74]

** The PATA linker was synthesized according to reference [49]



Compound 11.1.11: Coupling of PATA linker 11.1.10. Compound **11.1.9** (100 mg, 0.231 mmol) was dissolved in 3 mL dry DMF. Then, a solution of HATU (352 mg, 0.926 mmol) and DIPEA (0.161 mL, 0.926 mmol) in dry DMF was added and stirred for 10 min. Next, a solution of PATA (188 mg, 0.926 mmol) in dry DMF (4 mL) was added. The reaction mixture was stirred for 16 h at room temperature. The solvent was evaporated in vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.17) to give **11.1.11** as a white powder (151 mg, 82% yield). ^{1}H-NMR (500 MHz, CD_3OD); δ 7.8 (d, J = 6.86 Hz, 2 H), 7.65 (d, J = 6.66 Hz, 2 H), 7.35 (m, 4H), 7.2 (t, 2 H), 4.40 (m, 4 H), 4,00 (t, J = 6.5 Hz, 2 H), 3.65 (s, 1 H), 3.60 (s, 1 H) 2.35–2.10 (m, 2 H), 1,2 (s, 3 H), 0.90 (s, 3 H), 0.85 (d, J = 6.40 Hz, 3 H), 0.80 (s, 3 H). ^{13}C NMR (125 MHz, CD_3OD): δ = 175.93 (COOR-C24), 169.69 (CONH-C3), 169.16 (CONH-C12), 154.65 (2CONH linker), 142.85 (C), 142.76 (C), 135.49 (C), 135.01 (C), 131.03 (CH), 130.78 (CH), 129.83 (CH), 128.82 (CH), 128.69 (CH), 128.60 (CH), 128.54 (CH), 122.14

(CH), 78.09 (C), 76.37 (C), 67.04 (CH_2- Pr ester), 55.80 (CH-alkyne), 54.75 (CH-alkyne), 51.91 (CH), 50.62 (CH), 45.91 (C), 44.08 (CH), 43.85 (CH_2), 43.76 (CH_2), 37.32 (CH), 37.00 (CH_2), 36.02 (CH), 35.41 (C), 33.10 (CH_2), 32.10 (CH_2), 32.03 (CH_2), 28.50 (CH_2), 28.30 (CH_2), 28.16 (CH_2), 27.80 (CH_2), 26.80 (CH_2), 24.83 (CH_2), 23.89 (19-CH_3), 23.01 (CH_2), 18.69 (CH), 17.89 (CH), 17.26 (21-CH_3), 14.25 (CH), 13.14 (18-CH_3), 10.74 (CH_3). (ESI-MS m/z (% rel. int.) calcd. 802.4 (100), found 803.4 (100) [M+ H^+]. HR-MS (ES) m/z calcd. for $C_{49}H_{62}N_4O_6$ 802.4669, found 803.4734 (M + H^+; Δ = 1 ppm).



Compound 11.1.11

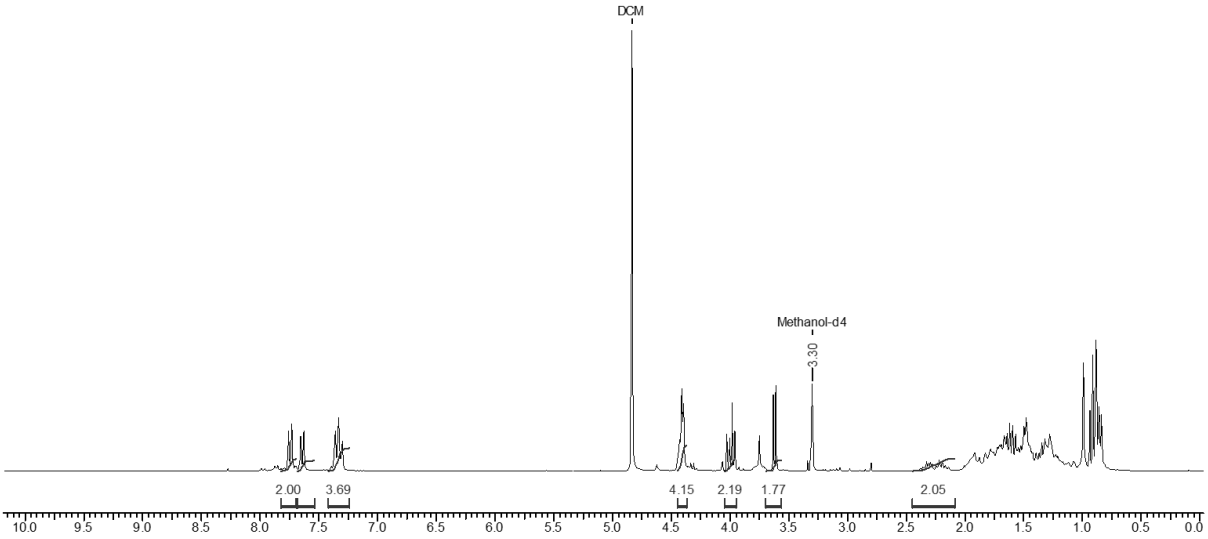


Figure 11.1.3. ^{1}H NMR spectrum of compound **11.1.11**.

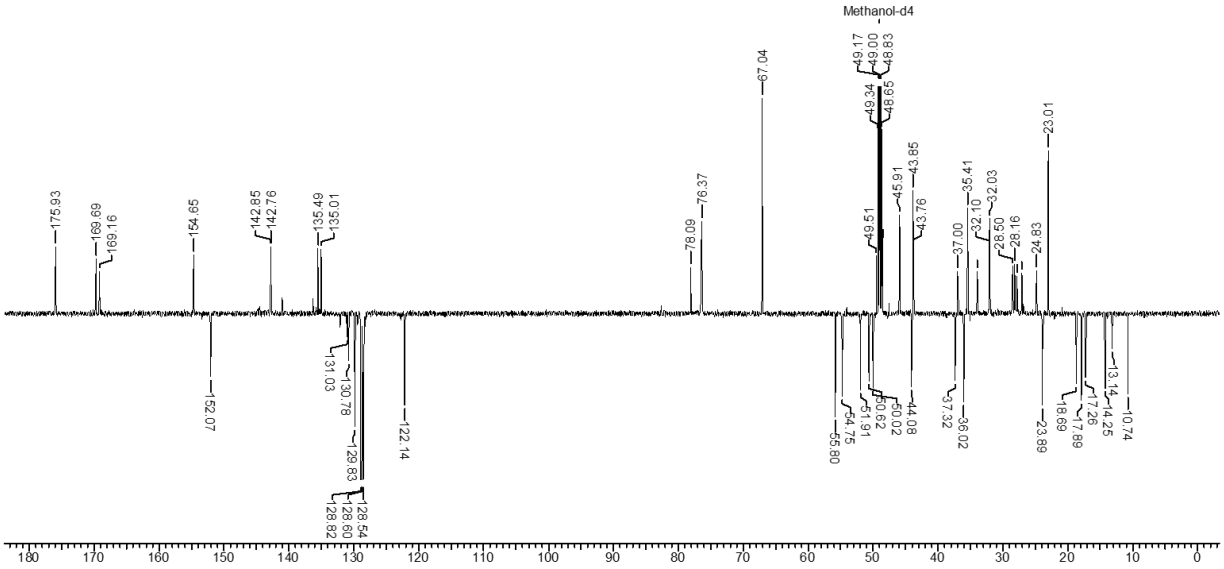
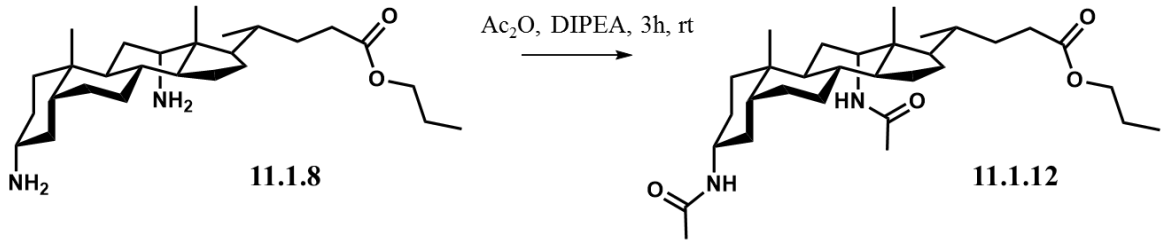


Figure 11.1.4. ^{13}C NMR spectrum of compound **11.1.11**.

Diamino scaffold 11.1.8 capped with Ac_2O

Scaffold **11.1.8** (20 mg, 0.046 mmol) was dissolved in dry DCM (2 mL), then acetic anhydride (0.017 mL, 0.185 mmol) and DIPEA (0.064 mL, 0.37 mmol) were added. The reaction mixture was stirred at room temperature for 3h. Solvent was evaporated and compound **11.1.8** was analyzed without prior purification. (21.4 mg, 90 % yield). ^1H-NMR (500 MHz, CD_3OD); δ 4.25 (m, 4 H), 4.00 (t, J = 6.5 Hz, 2 H), 3.65 (br. m, 1 H), 2.31–2.40 (m, 2 H), 2.18–2.26 (m, 2 H), 2.1 (s, 3H), 1.90 (s, 3H), 1.2 (s, 3 H), 0.90 (s, 3 H), 0.85 (d, J = 6.40 Hz, 3 H), 0.80 (s, 3 H). (ESI-MS m/z (% rel. int.) calcd. 516.4 (100), found 517.4 (100) [M + H^+]. HR-MS (ES) m/z calcd. for $C_{31}H_{52}N_2O_4$ 516.3927, found 517.4008 (M + H^+; Δ = 2.6 ppm).



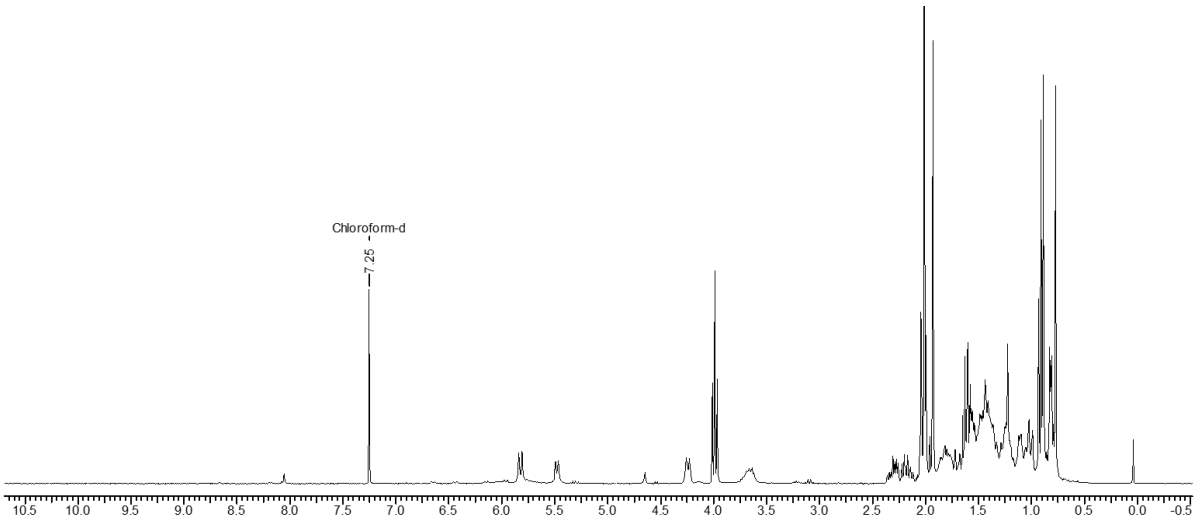
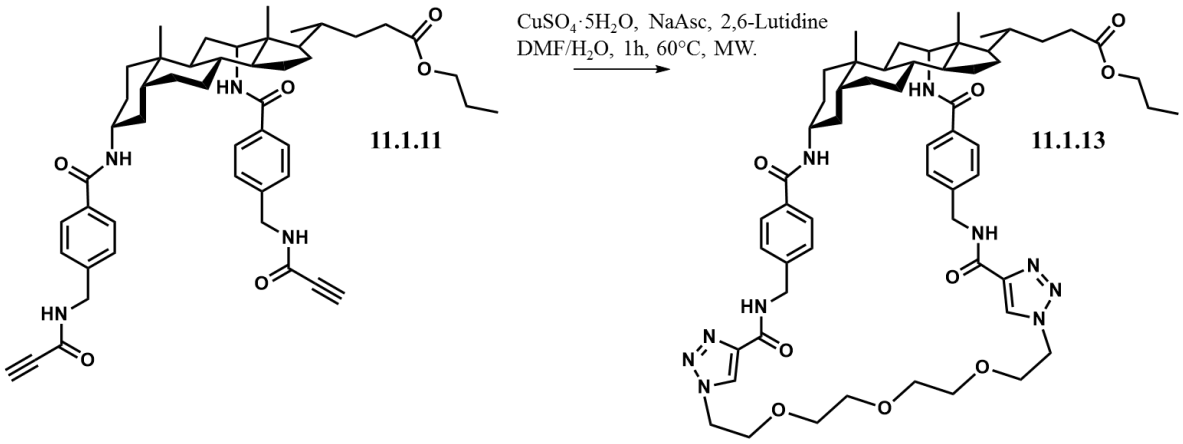


Figure 11.1.5. ^{1}H NMR spectrum of compound **11.1.12**.

Click cyclic receptor



Scaffold **11.1.11** (20 mg, 0.0249 mmol) was dissolved in 1 mL of dry DMF. 1.11-diazido-3,6,9, trioxaundecane (6.08 mg, 0.0249 mmol) was added followed by the addition of $CuSO_4 \cdot 5H_2O$ (3.1 mg, 0.012 mmol) in 0.1 mL dry DMF and 2,6-Lutidine (5.8 µL, 0.049 mmol). A solution of Na ascorbate (9.82 mg, 0.049 mmol) in 0.2 mL H_2O was subsequently added and the reaction mixture was stirred for 1h at 60°C under MW irradiation.

The solvent was evaporated and the reaction was extracted with DCM and a saturated solution of EDTA (3x40 mL). The organic phases were dried under $MgSO_4$ and filtered. The solvent was evaporated and the crude compound was purified by RP-HPLC. (ESI-MS m/z (% rel. int.) calcd. 1046.5 (100), found 1047.5 (100) [M+ H^+]. HR-MS (ES) m/z calcd. for $C_{57}H_{78}N_{10}O_9$ 1046.5953, found 1047.6032 (M + H^+; Δ = 0.6 ppm)..

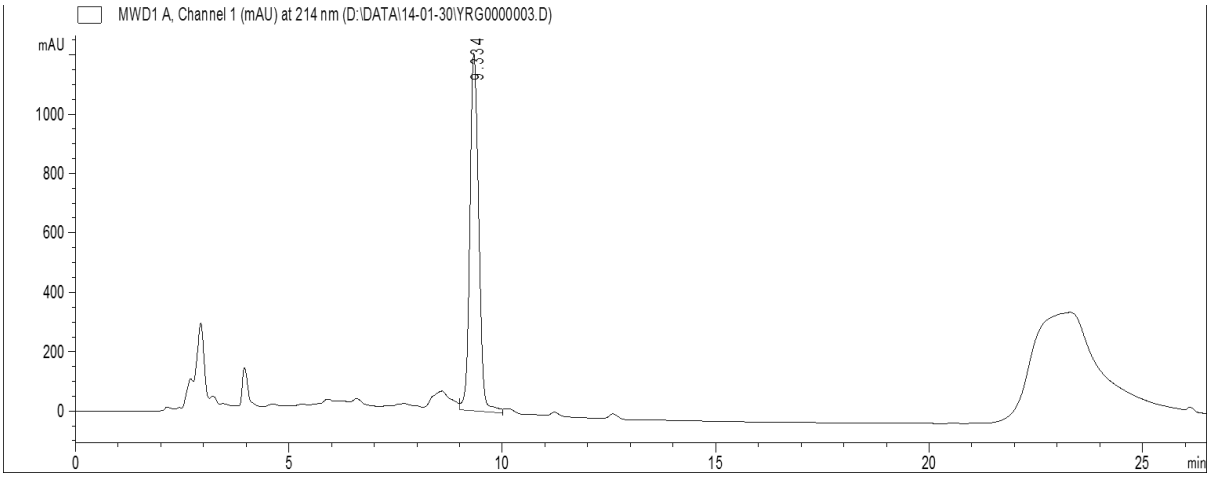


Figure 11.1.6: RP-HPLC Chromatogram of HPLC crude compound **11.1.13** (C18 100 Å column (250 x 10 mm, 5 µm, at 35 °C) using a gradient from 75 to 100 % CH_3CN in 15 minutes).

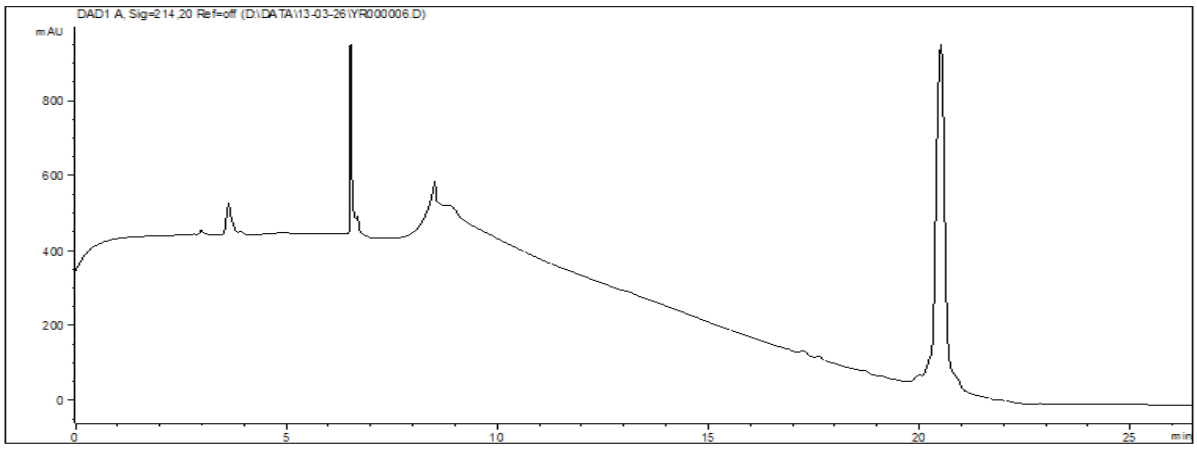


Figure 11.1.7: RP-HPLC Chromatogram of HPLC purified compound **11.1.13** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

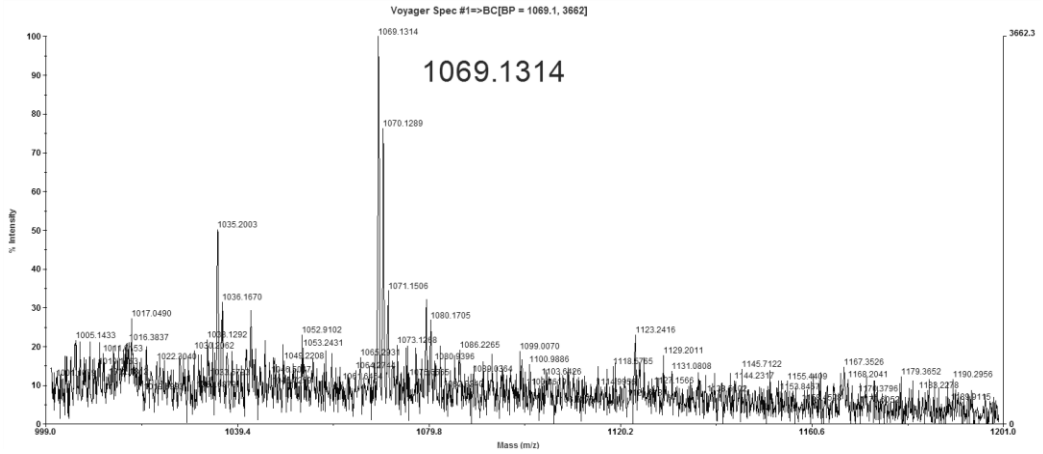


Figure 11.1.8: MALDI-TOF Spectrum of HPLC purified compound **11.1.13**. Calculated m/z (100%) = 1046.60; m/z (62.9%) = 1047.60. Found: 1069.13 = M + Na^+.

The synthesis of receptor **11.1.13** was more challenging due to the possibility of the diazido linker to react with the alkynes and form acyclic structure. In order to verify the obtention of **11.1.13** in the crude reaction mixture, we compared the analysis data with the results previously acquired from the synthesis of receptor **11.1.13** on TentaGel solid phase, on which cyclization could be better controlled and monitored. The first step was the incorporation of the photolabile linker for UV cleavage and a UV absorption moiety (tag) on the solid phase. The next step was the attachment of an orthogonally protected amine scaffold **11.1.13d**.

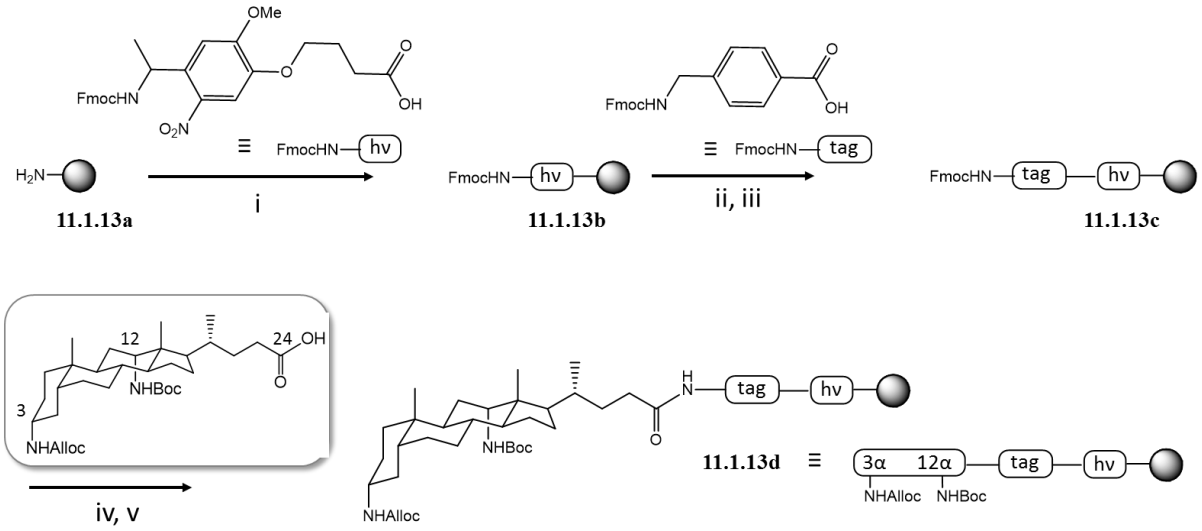


Figure 11.1.9: Synthesis scheme for the attachment of the steroid scaffold to the solid phase. i) PyBOP, DIPEA, DMF, 3h, rt. ii) 20% piperidine/DMF 2 x 30 min. iii) PyBOP, DIPEA, DMF, 3h, rt. iv) 20% piperidine/DMF 2, 5, 15 min. v) HATU, DIPEA, DMF, 2 x 4h rt.

Due to the orthogonal protection of the amines, selective coupling of PATA linker to C12 was achieved by the deprotection of the Boc. Further appendage of PATA linker afforded one alkyne for the conjugation with one azide of the diazo moiety via CuAAC. The Alloc deprotection of the second amine and the attachment of the PATA linker provided the acyclic receptor **11.1.13h** after UV cleavage. In order to obtain the cyclic form, compound **11.1.13g** was subjected to CuAAC to get compound **11.1.13i**, followed by UV irradiation to get the cyclic receptor **11.1.13j** cleaved from the resin. The comparison between the RP-HPLC profile of **11.1.13h** and **11.1.13j** with **11.1.13** allowed its structural characterization.

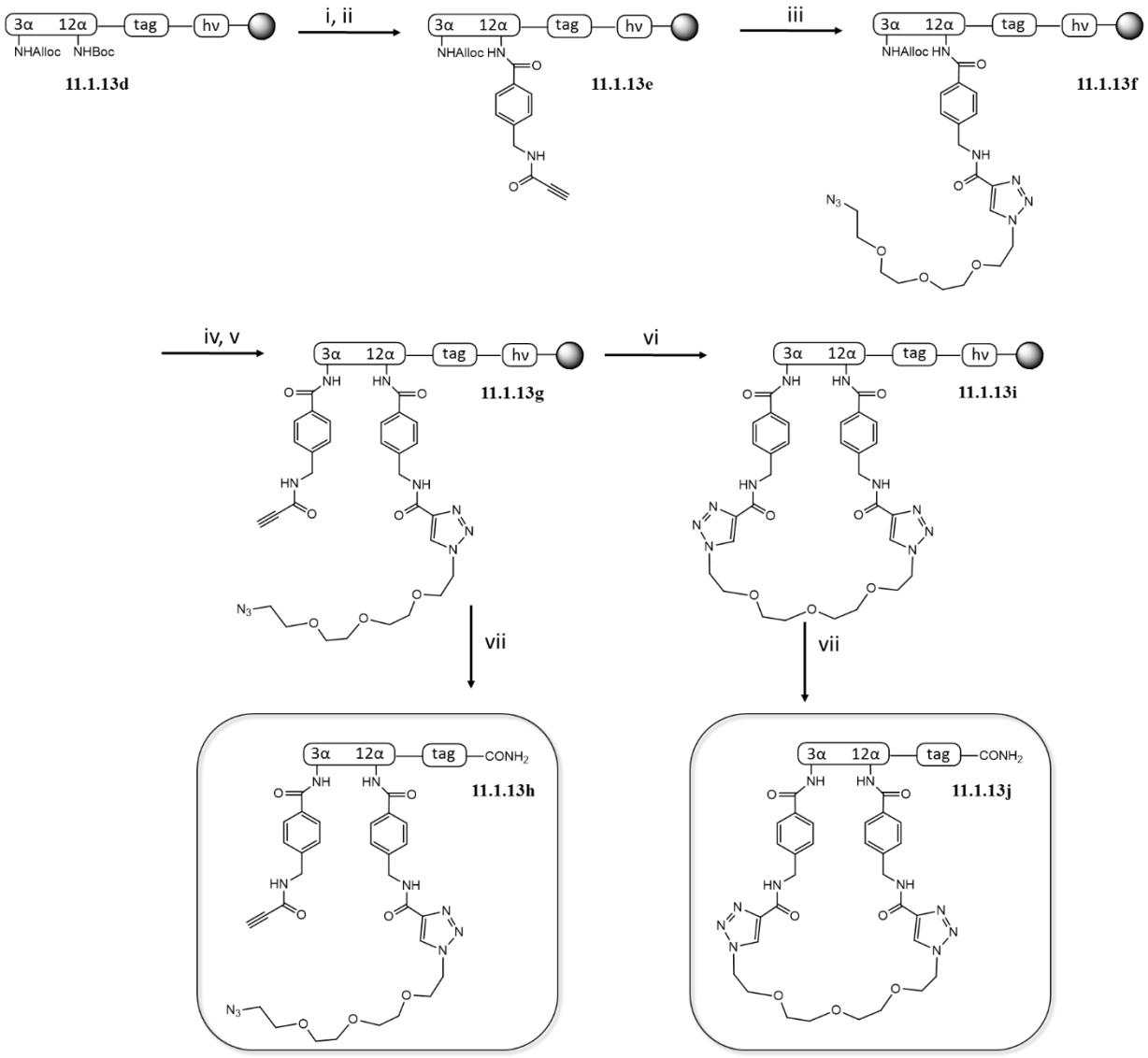
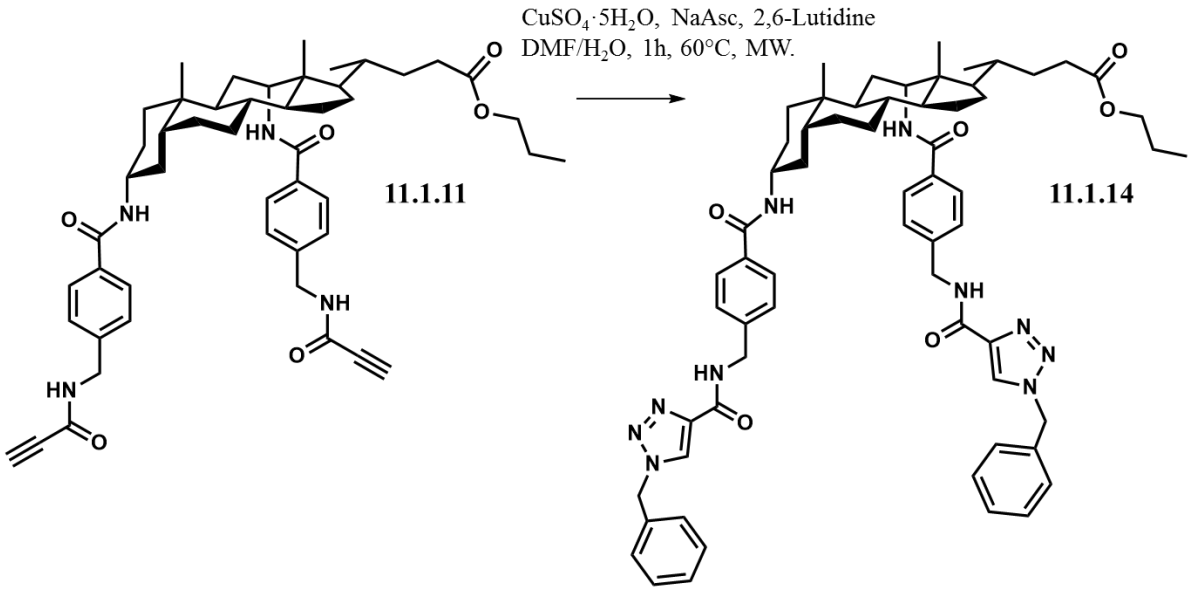


Figure 11.1.10: Synthesis scheme for cyclic and acyclic version of receptor **11.1.13** on solid phase. i) 20%TFA in DCM 2h rt. ii) PATA, HATU, DIPEA, DMF, 4h rt. iii) $CuSO_4 \cdot 5H_2O$, NaAsc, DIPEA, MW 70°C 1h. iv) Alloc deprotection. $Pd(PPh_3)_4$, Bu_3SnH, morpholine, DCM 2h rt. v) PATA, HATU, DIPEA, DMF, 4h rt. vi) $CuSO_4 \cdot 5H_2O$, NaAsc, DIPEA, MW 70°C 1h. vii) UV irradiation, 3h.

Click acyclic receptor



Scaffold **11.1.11** (20 mg, 0.0249 mmol) was dissolved in 1 mL of dry DMF. Benzyl azide (6.52 mg, 0.049 mmol) was then added followed by the addition of $CuSO_4$•$5H_2O$ (3.1 mg, 0.012 mmol) in 0.1 mL dry DMF and 2, 6-Lutidine (5.8 µL, 0.049 mmol). A solution of Na ascorbate (9.82 mg, 0.049 mmol) in 0.2 mL H_2O was subsequently added and the reaction mixture was stirred for 1h at 60°C under MW irradiation.

The solvent was evaporated and the reaction was extracted with DCM and a saturated solution of EDTA (3x40 mL). The organic phases were dried under $MgSO_4$ and filtered. The solvent was evaporated and the crude compound was purified by RP-HPLC. ESI-MS m/z (% rel. int.) calcd. 1068.5 (100), found 1069.5 (100) [M+ H^+]. HR-MS (ES) m/z calcd. for $C_{63}H_{76}N_{10}O_6$ + H^+ 1068.5949, found 1069.6027 (M + H^+; Δ = 0.5 ppm).

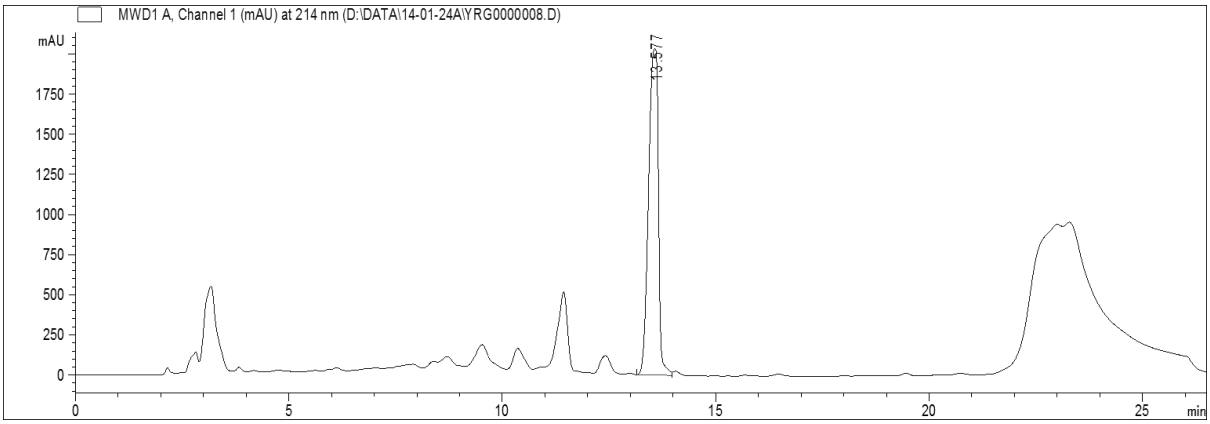


Figure 11.1.11: RP-HPLC Chromatogram of HPLC crude compound **11.1.15** (C18 100 Å column (250 x 10 mm, 5 µm, at 35 °C) using a gradient from 75 to 100 % CH_3CN in 15 minutes).

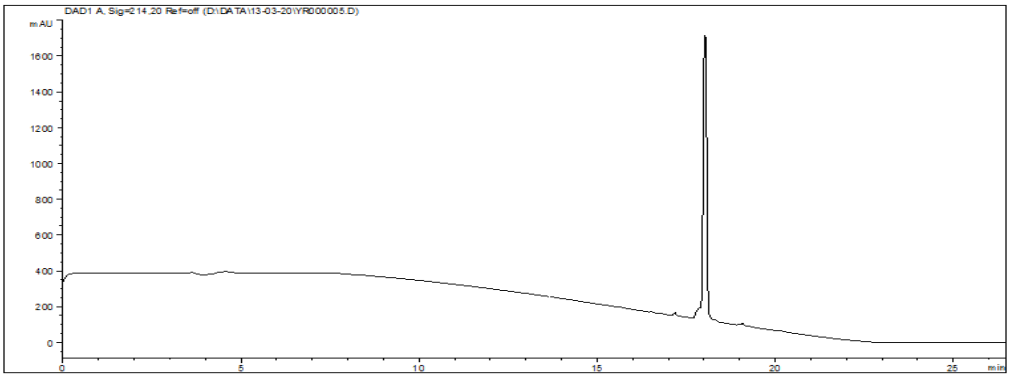


Figure 11.1.12: RP-HPLC Chromatogram of HPLC purified compound **11.1.14** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

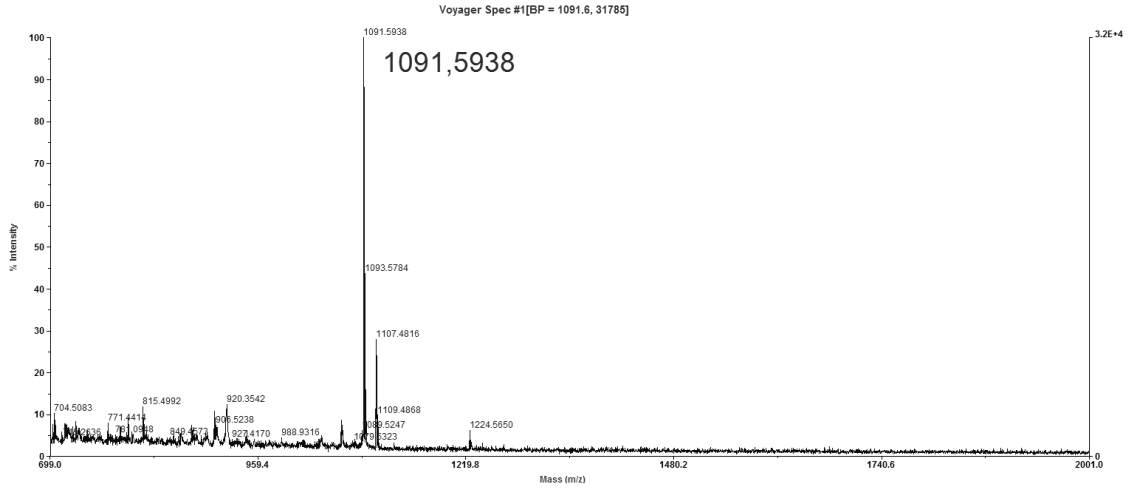
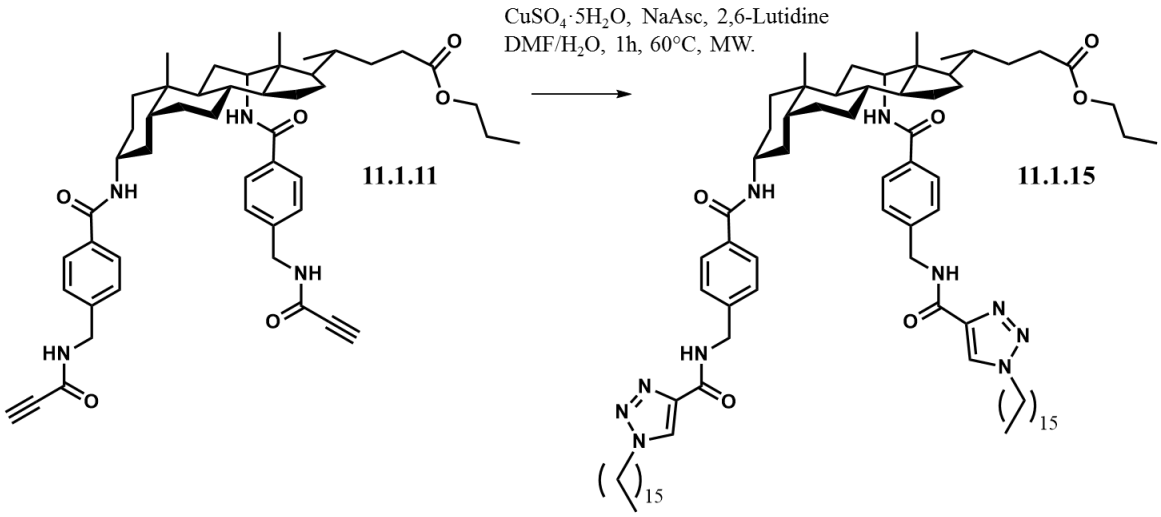


Figure 11.1.13: MALDI-TOF Spectrum of crude compound **11.1.14**. Calculated m/z (100%) = 1068.59; m/z (69.2%) = 1069.60. Found: 1091.59 = M + Na^+.

Click aliphatic receptor



Scaffold **11.1.11** (20 mg, 0.0249 mmol) was dissolved in 1 mL of dry DMF. Hexadecylazide **11.1.2** (13.3 mg, 0.049 mmol) was then added followed by the addition of $CuSO_4 \cdot 5H_2O$ (3.1 mg, 0.012 mmol) in 0.1 mL dry DMF and 2,6-Lutidine (5.8 µL, 0.049 mmol). A solution of Na ascorbate (9.82 mg, 0.049 mmol) in 0.2 mL H_2O was subsequently added and the reaction mixture was stirred for 1h at 60°C under MW irradiation.

The solvent was evaporated and the reaction was extracted with DCM and a saturated solution of EDTA (3x40 mL). The organic phases were dried under $MgSO_4$ and filtered. The solvent was evaporated and the residue was purified by column chromatography (silica gel DCM/MeOH 9:1) (62 % yield). ESI-MS m/z (% rel. int.) calcd. 1337.0 (100), found 1338.0 (100) [M+ H^+]. HR-MS (ES) m/z calcd. for $C_{63}H_{76}N_{10}O_6$ 1337.00183, found 1396.0153 (M + OAc^-; Δ = 0.3 ppm)..

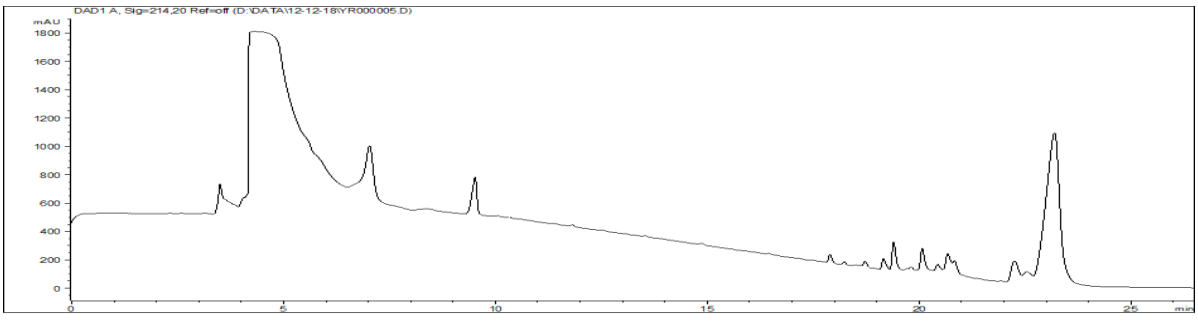


Figure 11.1.14: RP-HPLC Chromatogram of HPLC crude compound **11.1.15** (C4, 300Å column using a gradient from 50 to 100 % CH_3CN in 15 minutes).

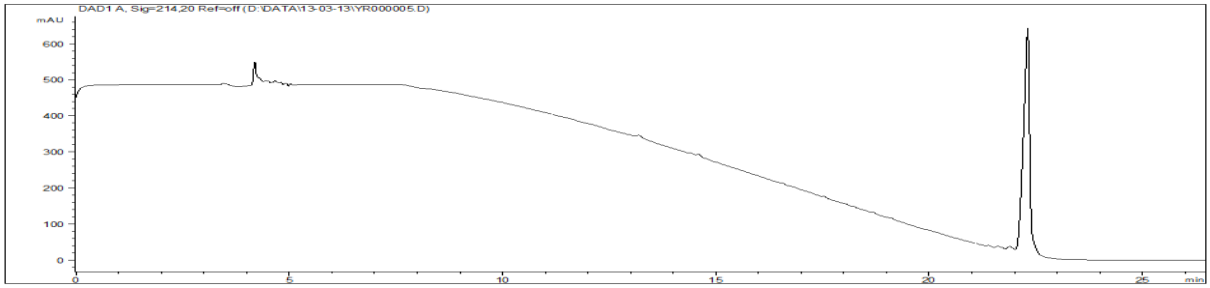


Figure 11.1.15: RP-HPLC Chromatogram of HPLC purified compound **11.1.15** (C4, 300Å column using a gradient from 50 to 100 % CH_3CN in 15 minutes)

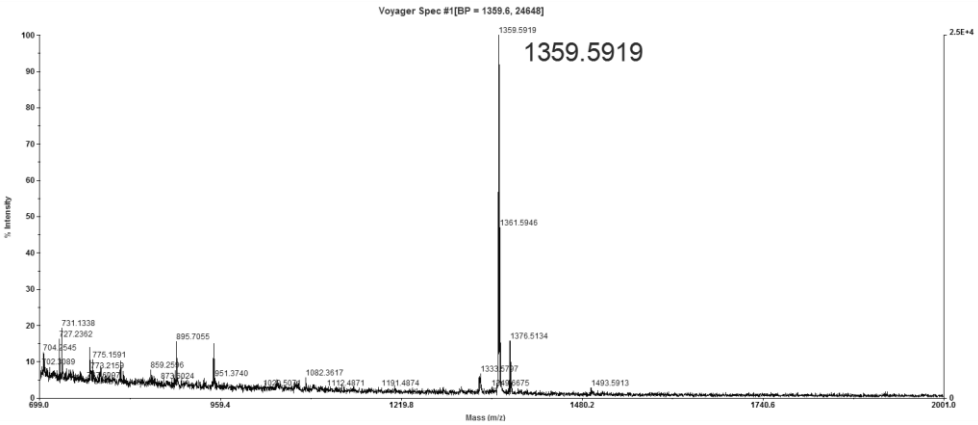


Figure 11.1.16: MALDI-TOF Spectrum of compound **11.1.15**. Calculated m/z (100%) = 1337.00; m/z (83.9%) = 1338.01. Found: 1359.59 = M + Na^+.

11.1.2 ITC measurements

Preparation of the solution of the ligand [phosphatidylserine (molecular weight: 790 g/mol)]

Concentration: 2400 µM: Dissolution of 3.792 mg in 2 mL of $CHCl_3$.

General protocol

Titrations were performed by introducing 1 mL of the solution of the receptor at the desired concentration in the measurement cell and 250 µL solution of phosphatidylserine (ligand) at the required concentration was loaded in the syringe. Titrations consisted in one injection of 5 µl followed by 24 injections of 10 µl at 300 s time intervals in the cell with stirring at 300 rpm.

Blank

In order to consider the influence of the solvent on the heat exchange during the titrations with the receptors, a blank experiment was performed. The procedure consisted on the injection of solvent from the syringe to the sample cell, in which only the solvent was present.

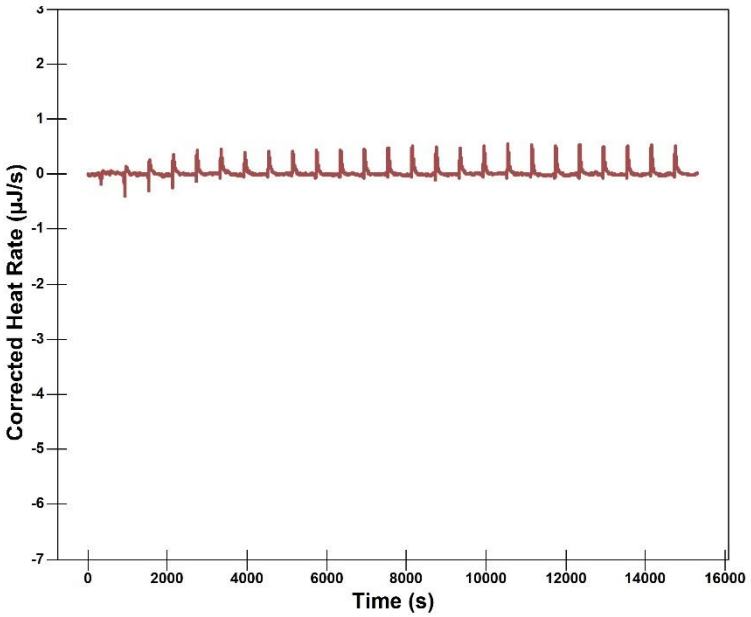


Figure 11.1.17: ITC curve of the titration with the solvent as blank.

Receptor 11.1.12

A 240 µM solution of the receptor **11.1.12** was prepared by weighing 1.24 mg in 10 mL of $CHCl_3$. The titration was performed in the same conditions as described previously.

Conclusion: After correction for the dilution heat, no significant heat exchange corresponding to the formation of a specific interaction complex could be observed.

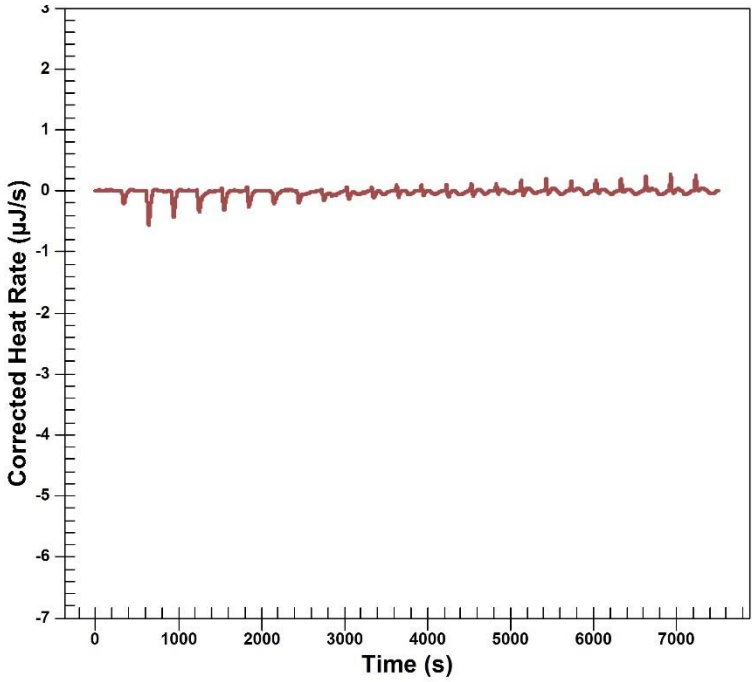


Figure 11.1.18: Receptor **11.1.12** at 240 µM with phosphatidylserine at 2400 µM.

Receptor 11.1.13

In order to get the 235 µM solution, we dissolved 2.1 mg of the product in 8.53 mL of $CHCl_3$. We ran the experiment with the concentrated solutions as previously described but no binding was observed.

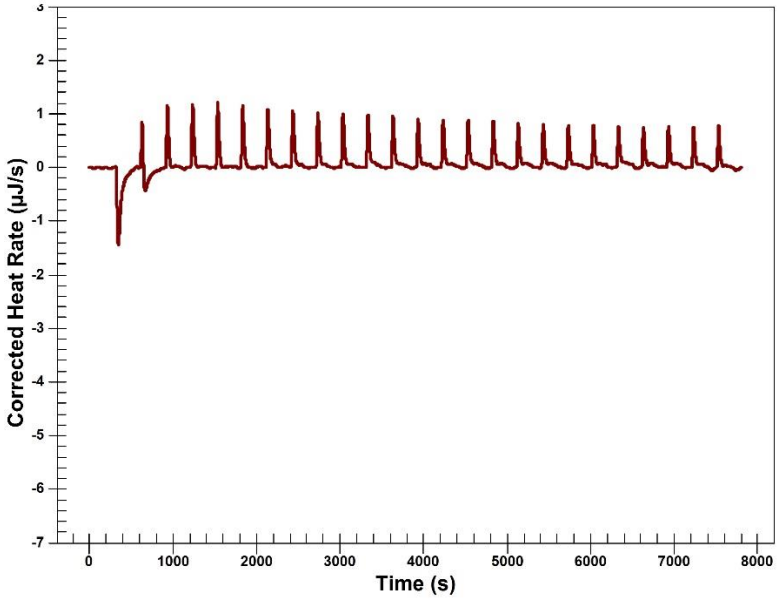


Figure 11.1.19: Receptor **11.1.13** at 235 µM with phosphatidylserine at 2400 µM. A small interaction is observed, however with a 0.08 stoichiometry, which probably do not correspond to a specific interaction with the receptor but to a non-specific interaction or to the presence of an impurity.

Receptor 11.1.14

In order to get the 240 µM solution, we dissolved 3.42 mg of the product in 13.32 mL of $CHCl_3$. We ran the same experiment previously described but no binding was observed.

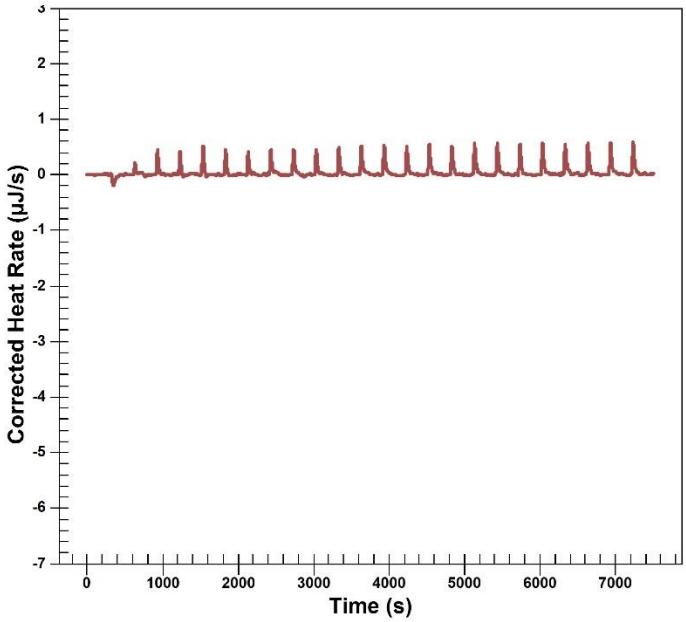


Figure 11.1.20: Receptor **11.1.14** at 240 µM with phosphatidylserine at 2400 µM.

Receptor 11.1.15

We ran the experiment with a 240 µM concentration of reception and 2400 µM concentration of ligand. We dissolved 3.46 mg of the product in 10.77 mL of $CHCl_3$.

Conclusion: Receptor **11.1.15** showed affinity for phosphatidylserine with Ka in the 10^5 range

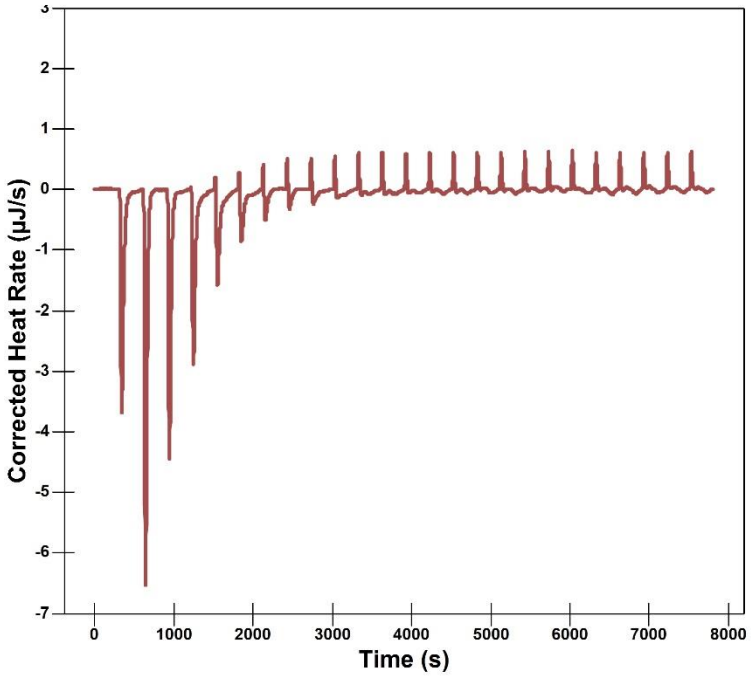


Figure 11.1.21: Receptor **11.1.15** at 240 µM with phosphatidylserine at 2400 µM. A binding isotherm was detected with a Ka of 1.4 10^5 M^{-1}, a ΔH° of 10.8 kJ/mol and a stoichiometry of 0.34.

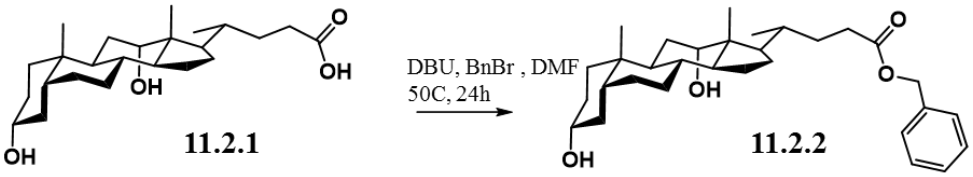
Summary of the studies of the interaction of the receptors with PS by ITC:

ITC measurements (ligand: phosphatidylserine) in $CHCl_3$				
Receptor	**Molecular weight (g/mol)**	**Conc. Receptor (µM)**	**Conc. Ligand (µM)**	**Binding/ K_a (M^{-1})**
11.1.12	516.756	240	2400	No
11.1.13	1047.291	235	2400	No
11.1.14	1069.341	240	2400	No
11.1.15	1337.946	240	2400	Yes/ 1.4 10^5

Chapter 11.2

11.2.1 Synthesis of Scaffolds

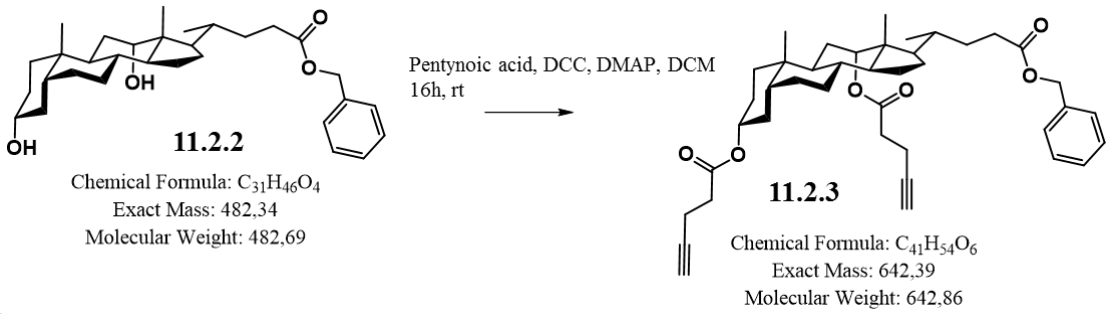
Benzyl 3α, 12α-dihydroxy-5β-cholan-24-oate (11.2.2). To a solution of deoxycholic acid (**11.2.1**) (3 g, 7.642 mmol) in DMF (7 mL), DBU (1.38 mL, 9.25 mmol) and benzyl bromide (1.1 mL, 9.25 mmol) were added and the mixture was stirred at 50 °C for 24 h. DMF was removed under vacuum, 10 % HCl (80 mL) was added to the residue and the product was extracted with CH_2Cl_2 (100 mL). The organic layer was washed with water (100 mL) and aq. $NaHCO_3$ (100 mL) and dried over anhyd. Na_2SO_4. The crude product obtained after removal of the solvent was purified using column chromatography on silica gel using 5 % EtOAc/$CHCl_3$ (Rf: 0.38) to yield 78 % of compound **11.2.2** as a white solid (2.81 g). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5 H), 5.15 (d, 1H, J = 14.13 Hz), 5.11 (d, 1H, J = 14.13 Hz), 3.92 (s, 1 H), 3.60 (br, m, 1 H), 2.45 – 2.36 (m, 1 H), 2.30 – 2.20 (m, 1 H), 0.90 (d, J = 6.28 Hz, 3 H), 0.80 (s, 3 H), 0.63 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 482.3 (100), found 541.3 (100) [M + OAc^-].



Chemical Formula: $C_{24}H_{40}O_4$
Exact Mass: 392,29
Molecular Weight: 392,57

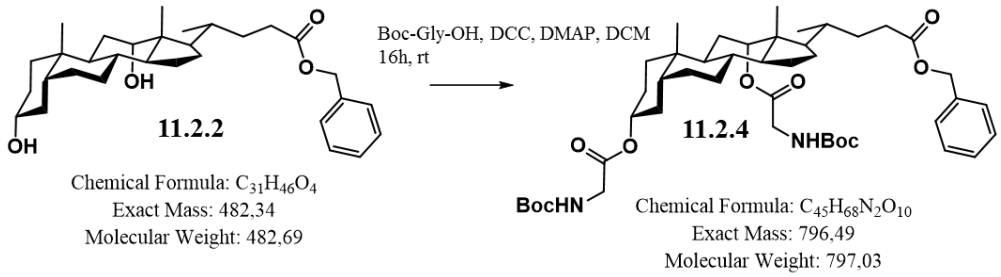
Chemical Formula: $C_{31}H_{46}O_4$
Exact Mass: 482,34
Molecular Weight: 482,69

Benzyl 3α, 12α-dipentynoate-5β-cholan-24-oate (11.2.3). Compound **11.2.2** (50 mg, 0.1037 mmol) was dissolved in 1.5 mL dry DCM. Then a solution of pentynoic acid (30.52 mg, 0.311 mmol), DCC (64.21 mg, 0.311 mmol) and DMAP (50.7 mg, 0.415 mmol) in dry DCM (2.5 mL) was added. The reaction mixture was stirred for 16h at room temperature. The precipitate formed was filtered off and the solvent was evaporated under vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.60) to give compound **11.2.3** as a yellow oil (64.6 mg, 97% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 5.15 (s, H), 5.15 (d, 1H, J = 13.56 Hz), 5.11 (d, 1H, J = 13.75 Hz), 4.6 (br, m, 1 H), 2.55-2.45 (m, 4 H), 2,43-2.40 (m, 4 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 173.90 (COOR-C24), 171.2 (COOR-C3), 170.87 (COOR-C12), 136.03 (C - benzyl group), 128.53 (2CH-benzyl ring), 128.25 (2CH- benzyl ring), 128.20 (CH-benzyl ring), 82.69 (C alkyne), 82.67 (C alkyne), 76.36 (CH ester), 74.57 (CH ester), 69.37 (CH_2), 68.95 (CH_2), 66.13 (CH_2 benzyl ester), 49.34 (CH), 47.61 (CH), 45.10 (C), 35.63 (CH), 34.80 (CH), 34.70 (CH), 34.69 (CH_2), 34.50 (CH), 34.45 (CH), 34.31 (CH), 34.05 (C), 33.91 (CH_2), 33.64 (CH_2), 32.16 (CH2), 31.24 (CH_2), 30.80 (CH_2), 27.34 (CH_2), 27.10 (CH_2), 26.83 (CH_2), 26.59 (CH_2), 23.44 (CH_2), 22.98 (19-CH_3), 17.50 (21 – CH_3), 14.42 (CH_2), 12.32 (18-CH3). ES-MS m/z (% rel. int.) calcd. 642.4 (100), found 660.4 (100) [M+ NH_4^+]. HR-MS (ES) m/z calcd. for $C_{41}H_{54}O_6$ 642.3920, found 660.4268 (M + NH_4^+; Δ = 1.2 ppm)..

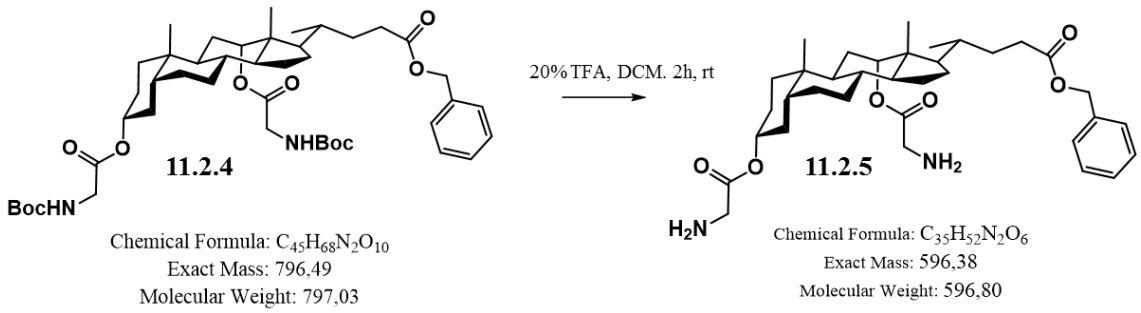


.

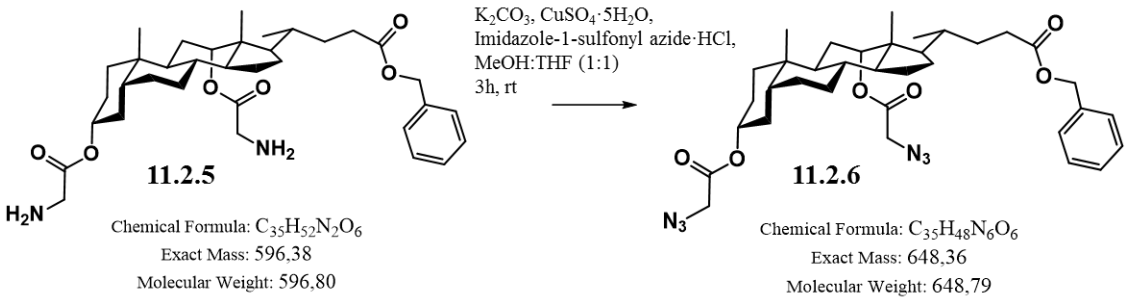
Benzyl 3α, 12α-Gly-5β-cholan-24-oate (11.2.4). Compound **11.2.3** (50 mg, 0.1037 mmol) was dissolved in 1.5 mL dry DCM. Then a solution of Boc protected glycine (54.52 mg, 0.311 mmol), DCC (64.21 mg, 0.311 mmol) and DMAP (50.7 mg, 0.415 mmol) in dry DCM (2.5 mL) were added. The reaction mixture was stirred for 16 h at room temperature. The precipitate formed was filtered off and the solvent is evaporated under vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.45) to give **11.2.4** as a yellow oil (76.8 mg, 93% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5 H), 5.2 (s, 1 H), 5.15 (m, 2 H), 4.6 (br, m, 1 H), 3.88 (d, 2 H), 3.85 (d, 2 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 1.40 (s, 18 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 796.5 (100), found 819.3 (100) [M+ Na^+].



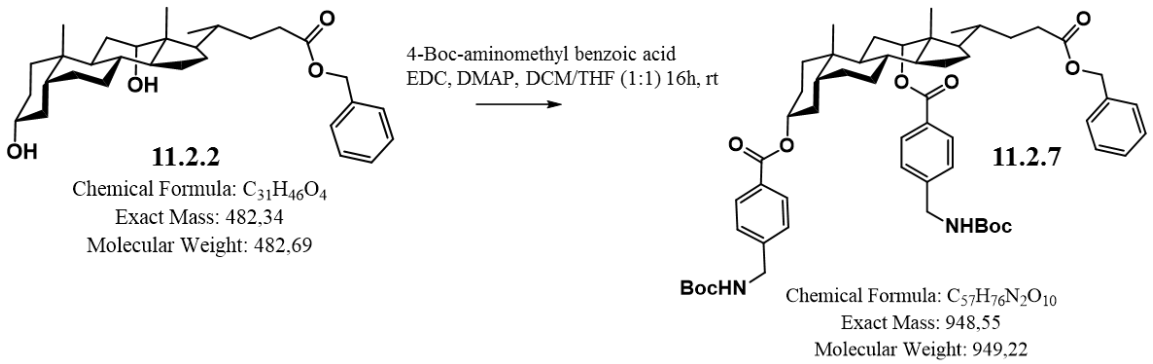
Compound 11.2.5 (Boc deprotection). Compound **11.2.4** (76.5 mg, 0.096 mmol) was treated with 10 mL of a solution of TFA (30%) in DCM for 2 h at room temperature. The solvent was evaporated and the white powder obtained was analyzed without further purification (Rf: 0.1, 56.7 mg, 99% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5 H), 5.2 (s, 1 H), 5.15 (m, 2 H), 4.6 (br, m, 1 H), 3.87 (m, 4 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 1.40 (s, 18 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 596.4 (100), found 597.3 (100) [M+ H^+].



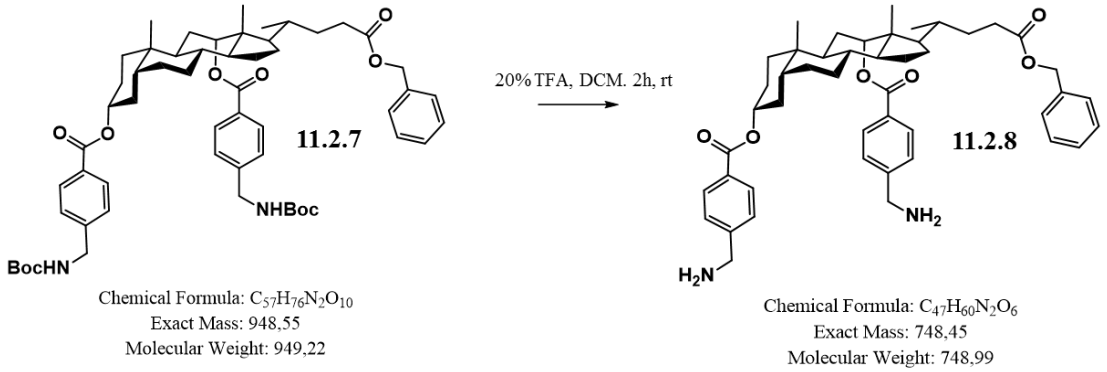
Compound 11.2.6 Diazotransfer reaction. Boc deprotected compound **11.2.5** (57 mg, 0.096 mmol) was dissolved in 2 mL MeOH/THF (1:1) and treated with K_2CO_3 (53 mg, 0.386 mmol) followed by $CuSO_4 \cdot 5H_2O$ (0.725 mg, 0.0029 mmol). To this mixture, imidazole-1-sulfonyl azide·HCl (61 mg, 0.290 mmol) dissolved in 1 mL MeOH/THF (1:1) was added to the reaction mixture and stirred at room temperature for 3 h under argon. The reaction was quenched with 4.5 mL of water and 0.5 mL of AcOH was added. Then, the reaction mixture was extracted 2 times with hexane (5 mL). The organic phase was dried under $MgSO_4$. The solvent was evaporated and the residue purified by column chromatography (silica gel DCM/MeOH 9.5:0.5; Rf: 0.20) to get a yellowish oil (59.7 mg, 96% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5 H), 5.2 (s, 1 H), 5.15 (d, 1H, J = 13.75 Hz), 5.10 (d, 1H, J = 13.94 Hz), 4.6 (br, m, 1 H), 3.98-3.90 (d, 2 H), 3.90-3.80 (d, 2 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 1.40 (s, 18 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 173.85 (COOR-C24), 167.76 (COOR-C3), 167.47 (COOR-C12), 135.99 (C - benzyl group), 128.52 (2CH-benzyl ring), 128.25 (2CH- benzyl ring), 128.19 (CH-benzyl ring), 77.97 (CH-C3), 75.97 (CH-C12), 66.14 (CH_2 benzyl ester), 50.91 (CH_2 ester glycine C3), 50.51 (CH_2 ester glycine C12), 47.64 (CH), 45.22 (CH), 45.06 (C), 41.67 (CH), 35.53 (CH), 33.95 (C), 31.93 (CH_2), 31.17 (CH_2), 30.70 (CH_2), 29.66 (CH_2), 27.24 (CH_2), 26.90 (CH_2), 26.70 (CH_2), 25.90 (CH_2), 25.77 (CH), 25.63 (CH), 25.40 (CH_2), 23.34 (CH_2), 22.85 (19-CH_3), 17.61 (21 – CH_3), 12.32 (18-CH_3). ES-MS m/z (% rel. int.) calcd. 648.4 (100), found 666.4 (100) [M+NH_4^+]. HR-MS (ES) m/z calcd. for $C_{35}H_{48}N_6O_6$ 648.3635, found 666.3968 (M + NH_4^+; Δ = 0.9 ppm).



Compound 11.2.7 (esterification with boc-aminomethylbenzoic acid). 4-Boc-aminomethyl benzoic acid (417.4 mg, 1.6596 mmol) was dissolved in 4 mL dry DCM/dry THF (1:1) at room temperature. EDC (0.294 mL, 1.6596 mmol) was added in portions and stirred for 30 min. DMAP (405.5 mg, 3.319 mmol) was added to the reaction mixture. Compound **11.2.2** (200 mg, 0.4149 mmol) dissolved in 6 mL dry DCM was then added and the reaction was stirred overnight. The solvent was evaporated and the residue purified by column chromatography (silica gel EtOAc/Hexane 3:1; Rf: 0.57) obtaining **11.2.7** with 82 % yield (322.7 mg). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2 H), 7.9 (d, J = 8.63 Hz, 2 H), 7.45 (d, J = 8,2 Hz, 2 H), 7.38 (d, J = 8 Hz, 2 H), 7.36 − 7.38 (m, 5 H), 5.3 (s, 1 H), 5.15 (m, 2 H), 4.7 (br, m, 1 H), 4.3 (br m, 4 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 2,00 (s, 18 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 948.5 (100), found 1007.5 (100) [M + OAc^-].



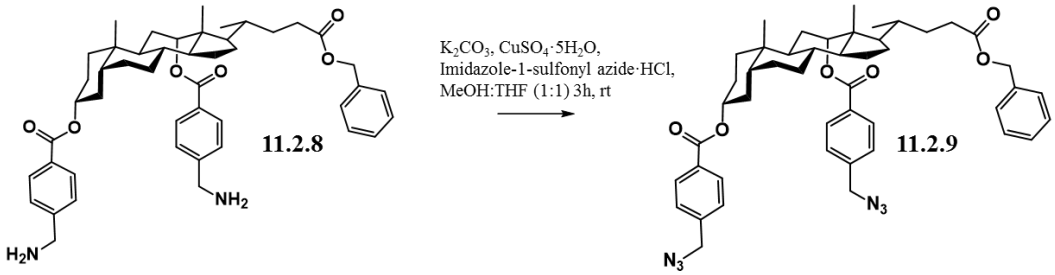
Compound 11.2.8 (Boc deprotection). Compound **11.2.7** (55 mg, 0.058 mmol) was treated with 10 mL of a solution of TFA (30%) in DCM for 2 h at room temperature. The solvent was evaporated and the white powder obtained was analyzed without further purification (Rf: 0.12, 43 mg, 99% yield. ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2 H), 7.9 (d, J = 8.63 Hz, 2 H), 7.45 (d, J = 8,2 Hz, 2 H), 7.38 (d, J = 8 Hz, 2 H), 7.36 − 7.38 (m, 5 H), 5.3 (s, 1 H), 5.15 (m, 2 H), 4.7 (br, m, 1 H), 4.3 (m, 2 H), 4.1-3.9 (m, 2 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 748.5 (100), found 749.5 (100) [M + H^+].



Compound 11.2.9. Diazotransfer reaction. Compound **11.2.8** (50 mg, 0.067 mmol) was dissolved in 2 mL MeOH/THF (1:1) and treated with K_2CO_3 (36.9 mg, 0.267 mmol) followed by $CuSO_4 \cdot 5H_2O$ (0.5 mg, 0.002 mmol). To this mixture, imidazole-1-sulfonyl azide·HCl (54 mg, 0.2 mmol) dissolved in 1 mL MeOH/THF (1:1) was added to the reaction mixture and stirred at room temperature for 3h under argon. The reaction was quenched with 4.5 mL of water and 0.5 mL of AcOH was added. Then, the reaction was extracted 2 times with hexane (5 mL). The organic phase was dried under $MgSO_4$. The solvent was evaporated and the residue purified by column chromatography (silica gel EtOAc/Hexane 1:1; Rf: 0.65) to obtain a yellow oil (51.5 mg, 96% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2 H), 7.9 (d, J = 8.63 Hz, 2 H), 7.45 (d, J = 8,2 Hz, 2 H), 7.38 (d, J = 8 Hz, 2 H), 7.36 − 7.38 (m, 5 H), 5.3 (s, 1 H), 5.18 (d, 1H, J = 13.75 Hz), 5.15 (d, 1H, J = 13.75 Hz), 4.7 (br, m, 1 H), 4.40 (s, 2 H), 4.38 (s, 2 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ^{13}C NMR (125 MHz, CD_3OD): δ = 173.88 (COOR-C24), 165.47 (COOR-C3), 165.40 (COOR-C12), 140.45 (C), 140.30 (C), 140.13 (C), 136.04 (C-benzyl ester), 130.76 (C), 130.01 (2CH), 129.97 (2CH),

128.52 (2CH), 128.22 (2CH), 128.17 (CH), 127.95 (2CH), 127.79 (2CH), 76.67 (CH ester linker C3), 74.83 (CH ester linker C12), 66.09 (CH_2 benzyl ester), 54.26 (CH_2 ester linker C3), 54.23 (CH_2 ester linker C12), 50.14 (CH), 48.00 (CH), 45.51 (C), 41.77 (CH), 35.80 (CH), 35.74 (CH), 34.72 (CH), 34.05 (C), 32.29 (CH_2), 31.24 (CH_2), 30.75 (CH_2), 30.10 (CH_2), 27.37 (CH_2), 26.82 (CH_2), 26.47 (CH_2), 25.99 (CH_2), 25.89 (CH_2), 23.50 (CH_2), 23.05 (19-CH_3), 17.46 (21-CH_3), 12.55 (18-CH_3). (ES-MS m/z (% rel. int.) calcd. 800.4 (100), found 818.4 (100) [M+ NH_4^+]. HR-MS (ES) m/z calcd. for $C_{47}H_{56}N_6O_6$ 800.42613, found 818.4592 (M + NH_4^+; Δ = 0.9 ppm).

.

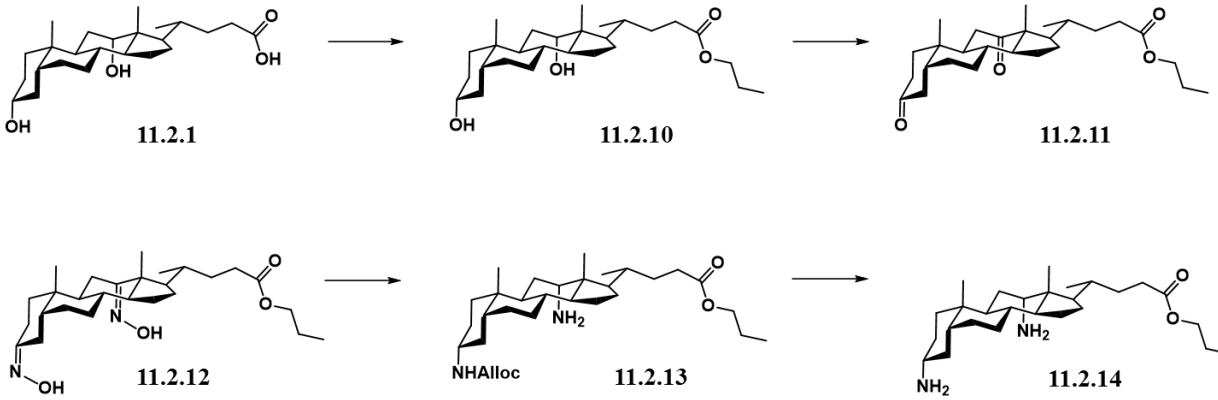


Chemical Formula: $C_{47}H_{60}N_2O_6$
Exact Mass: 748,45
Molecular Weight: 748,99

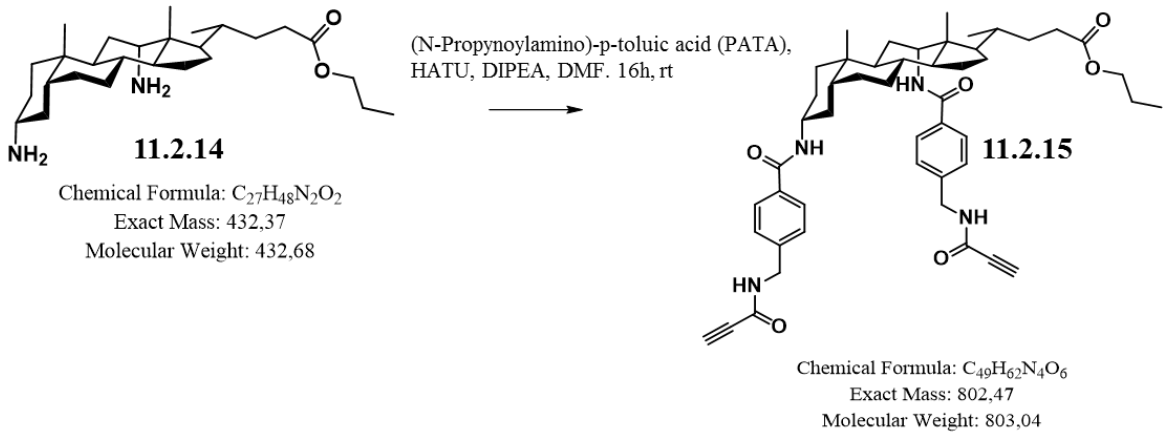
Chemical Formula: $C_{47}H_{56}N_6O_6$
Exact Mass: 800,43
Molecular Weight: 800,98

The Diamino scaffold 11.2.14 was synthesized according to reference 1 and 2 [48,74]

PATA linker was synthesized according to reference 3 [49]



Compound 15. Coupling of PATA linker. Refer to section 11.1.3 for synthesis and characterization of **11.2.15**



11.2.2 NMR Spectra for steroid template molecules:

Compound 11.2.2:

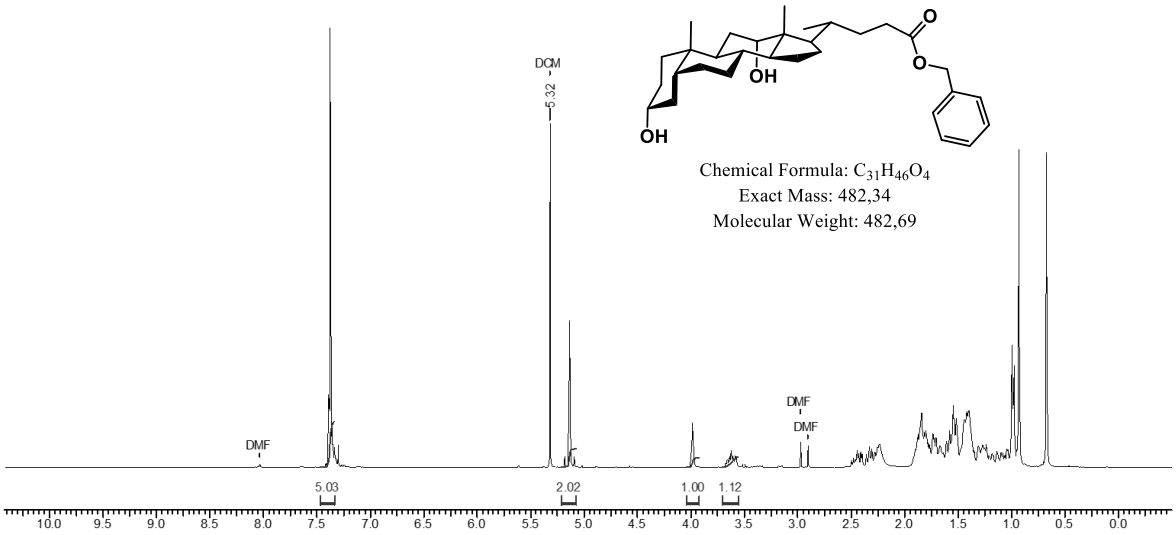


Figure 11.2.1. ^{1}H NMR spectrum of compound **11.2.2**.

Compound 11.2.3:

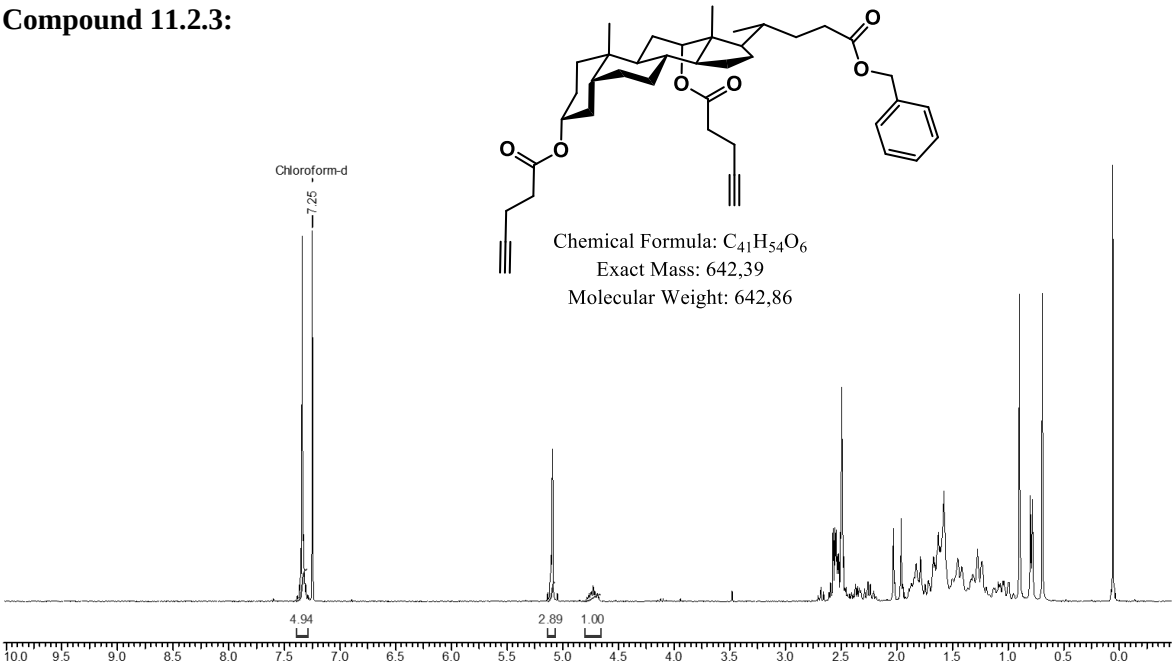


Figure 11.2.2. ^{1}H NMR spectrum of compound **11.2.3**.

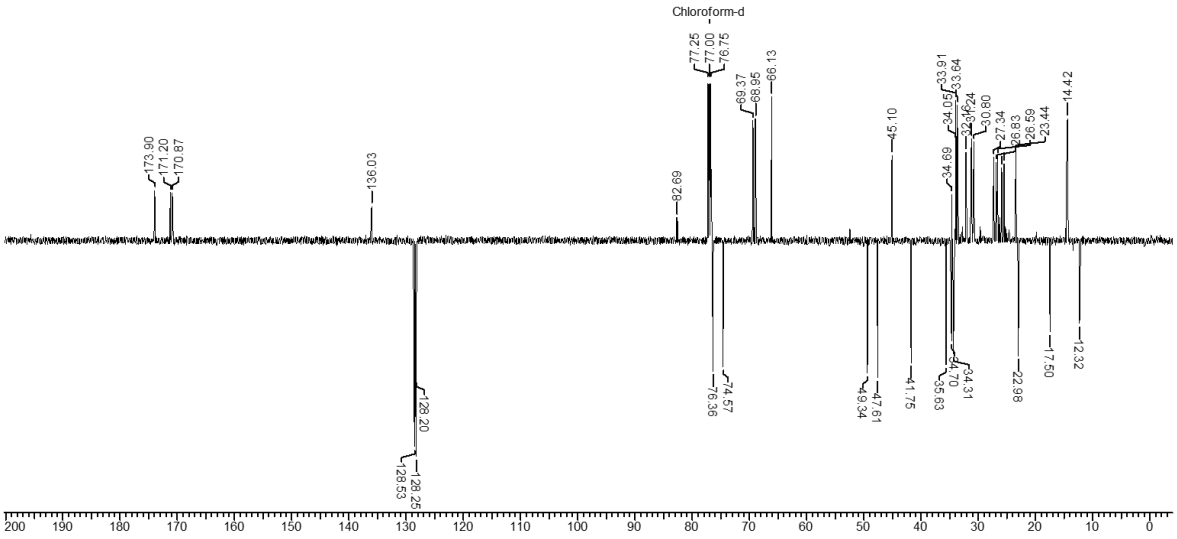


Figure 11.2.3. ^{13}C NMR spectrum of compound **11.2.3**.

Compound 11.2.4:

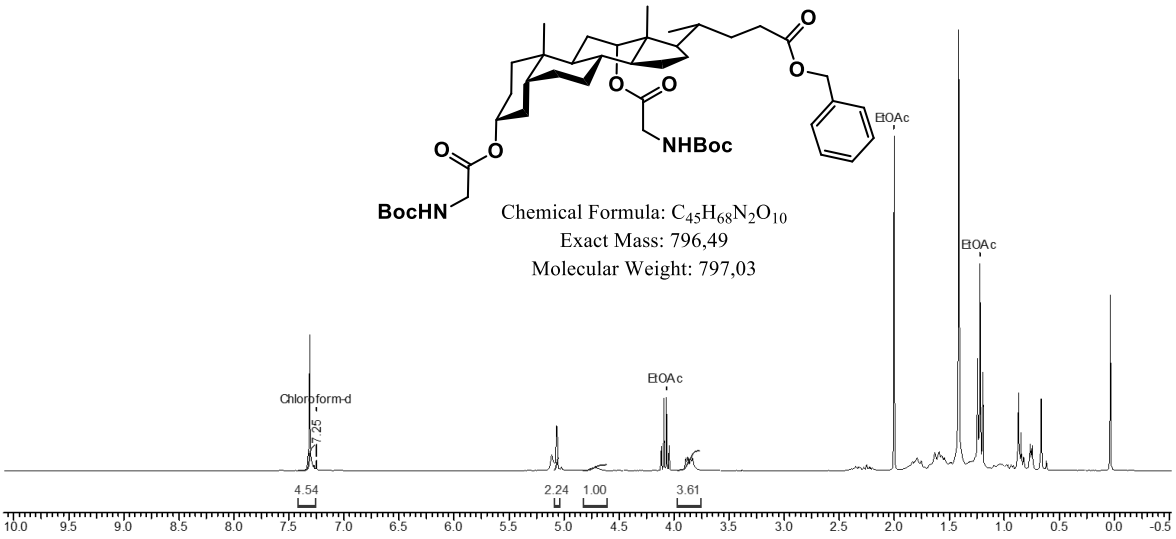


Figure 11.2.4. ^{1}H NMR spectrum of compound **11.2.4**.

Compound 11.2.5

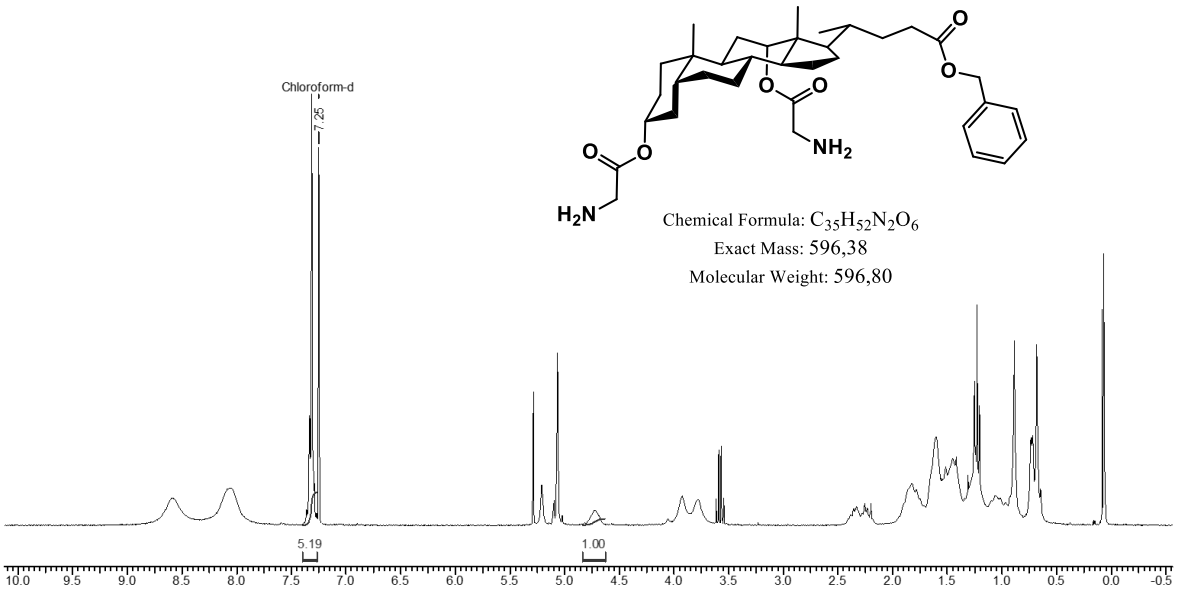


Figure 11.2.5. ^{1}H NMR spectrum of compound **11.2.5**.

Compound 11.2.6

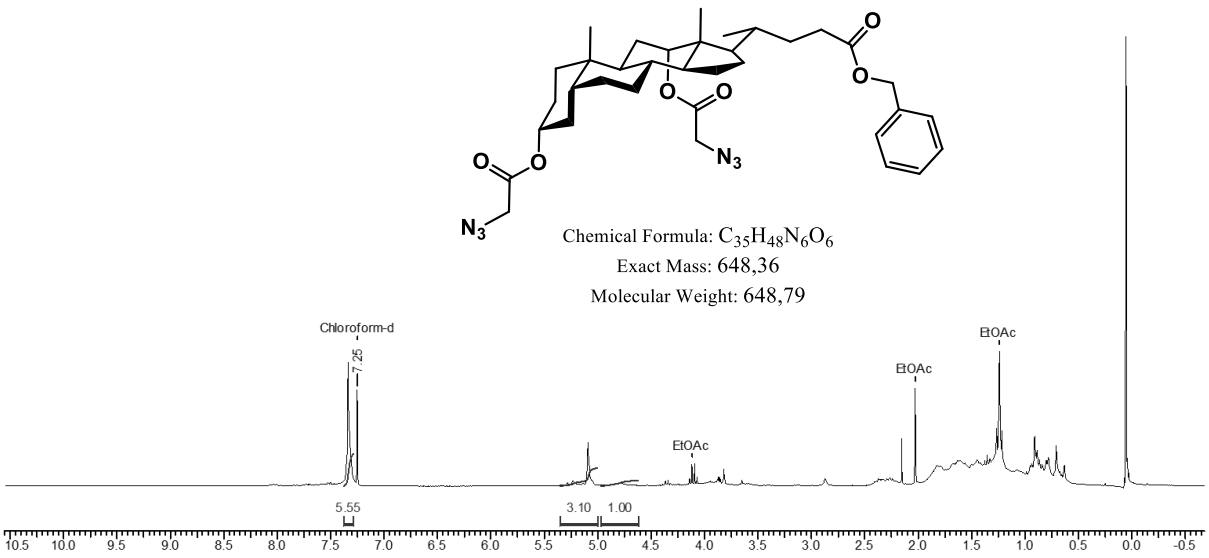


Figure 11.2.6. ^{1}H NMR spectrum of compound **11.2.6**.

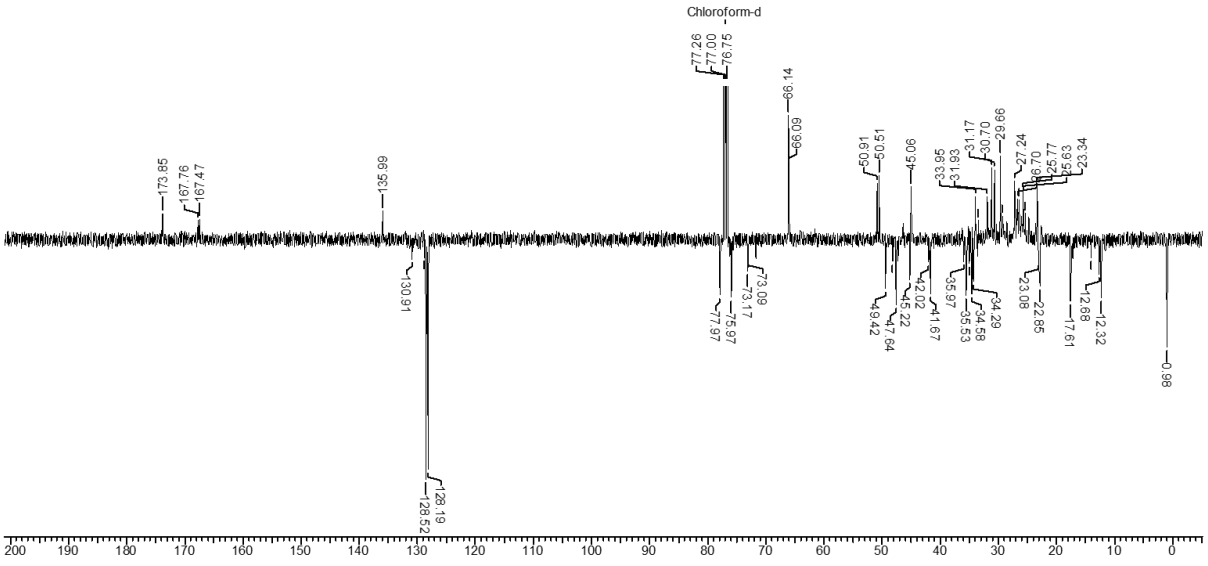


Figure 11.2.7. ^{13}C NMR spectrum of compound **11.2.6**.

Compound 11.2.7

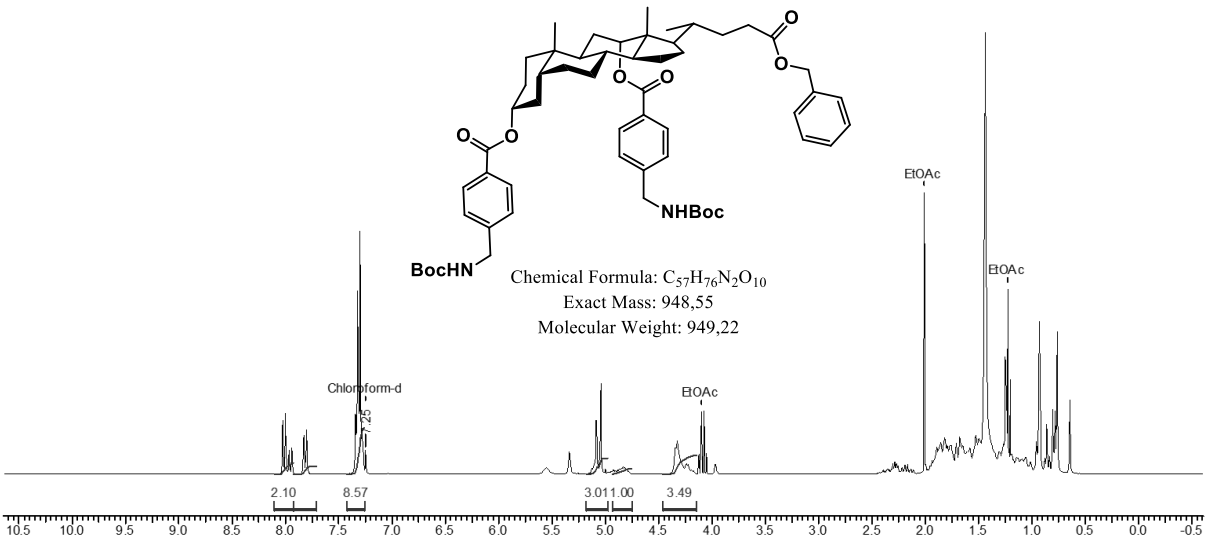


Figure 11.2.8. ^{1}H NMR spectrum of compound **11.2.7**.

Compound 11.2.8

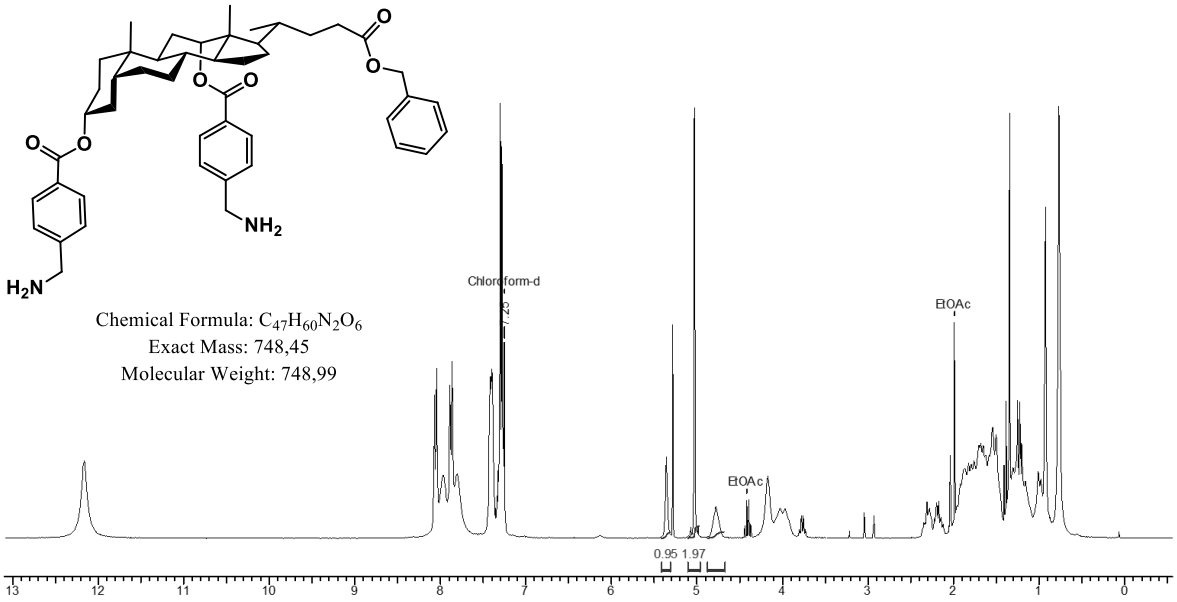


Figure 11.2.9. ^{1}H NMR spectrum of compound **11.2.8**.

Compound 11.2.9

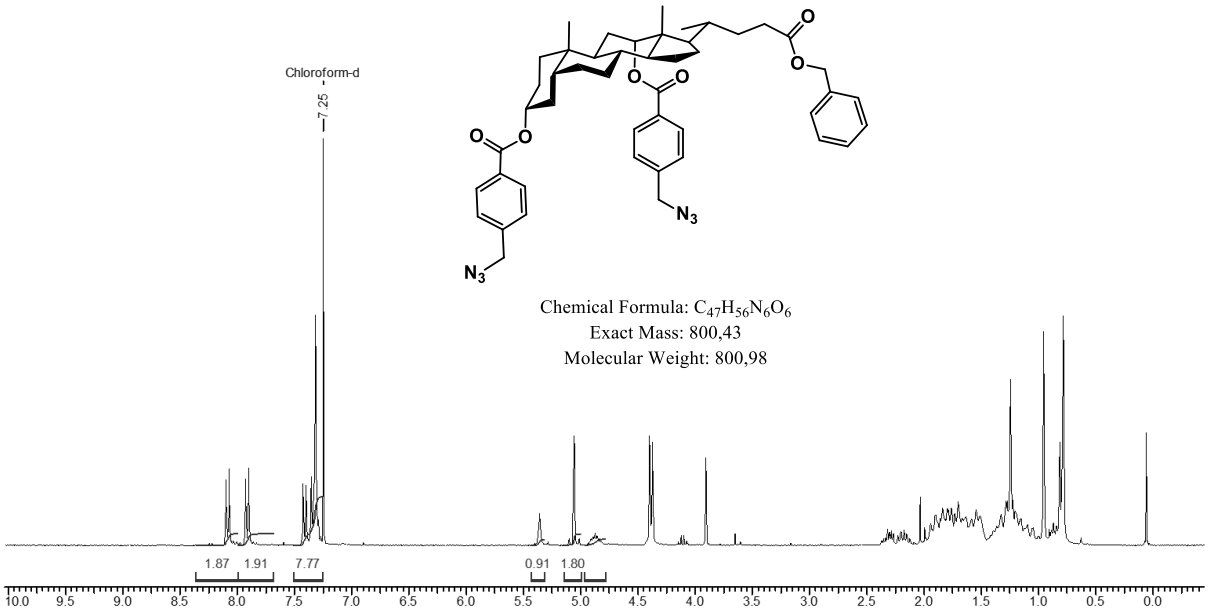


Figure 11.2.10. ^{1}H NMR spectrum of compound **11.2.9**.

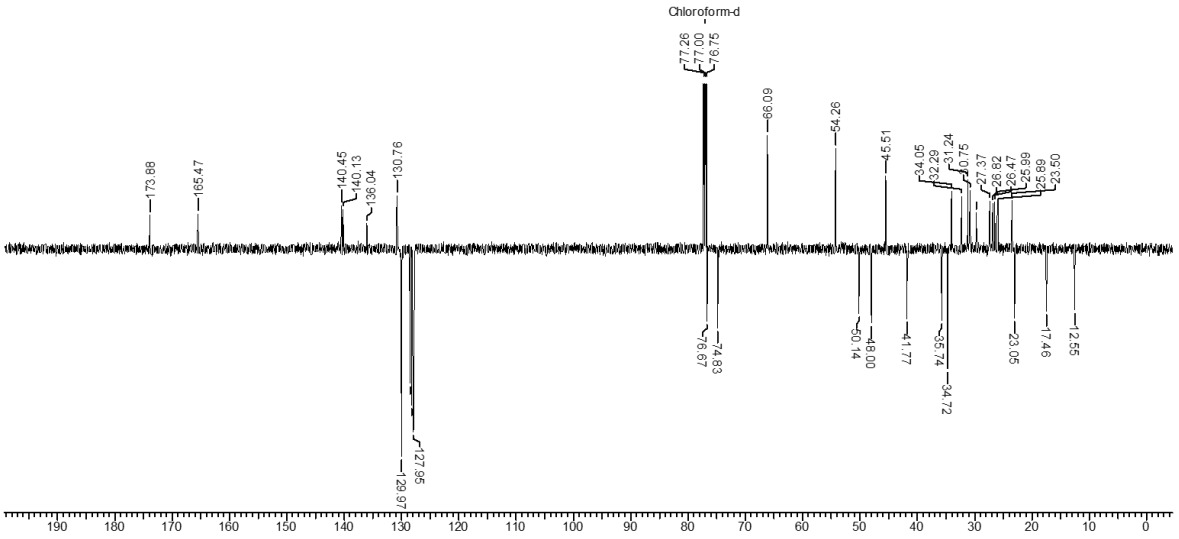


Figure 11.2.11. ^{13}C NMR spectrum of compound **11.2.9**.

Compound 11.2.15

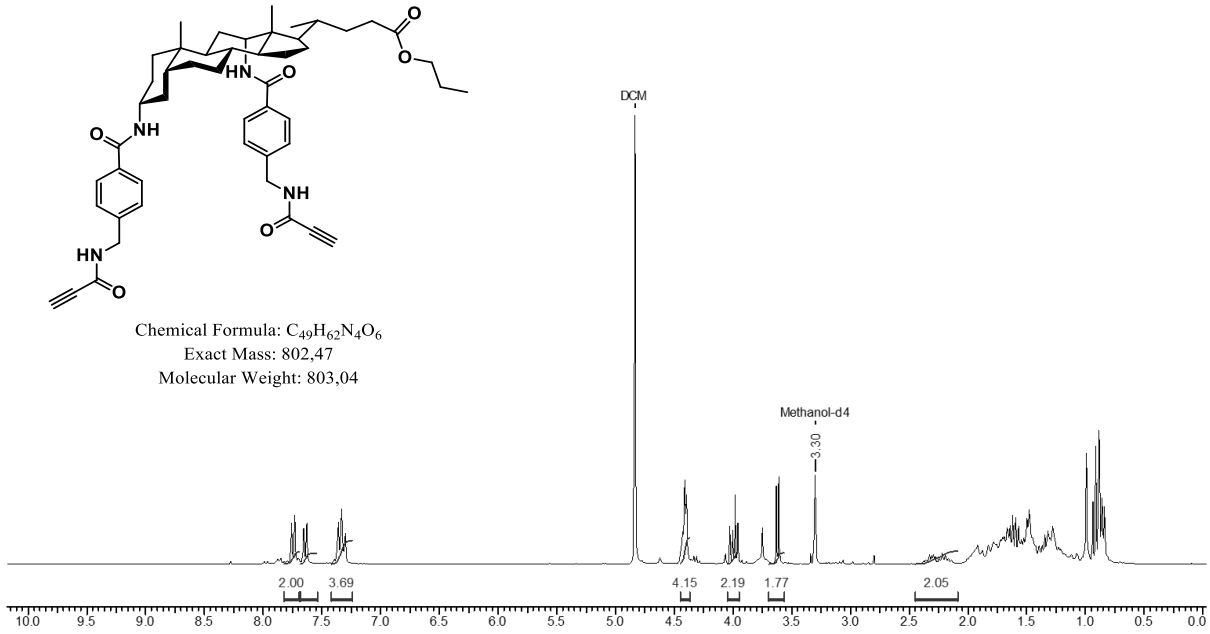


Figure 11.2.12. ^{1}H NMR spectrum of compound **11.2.15**.

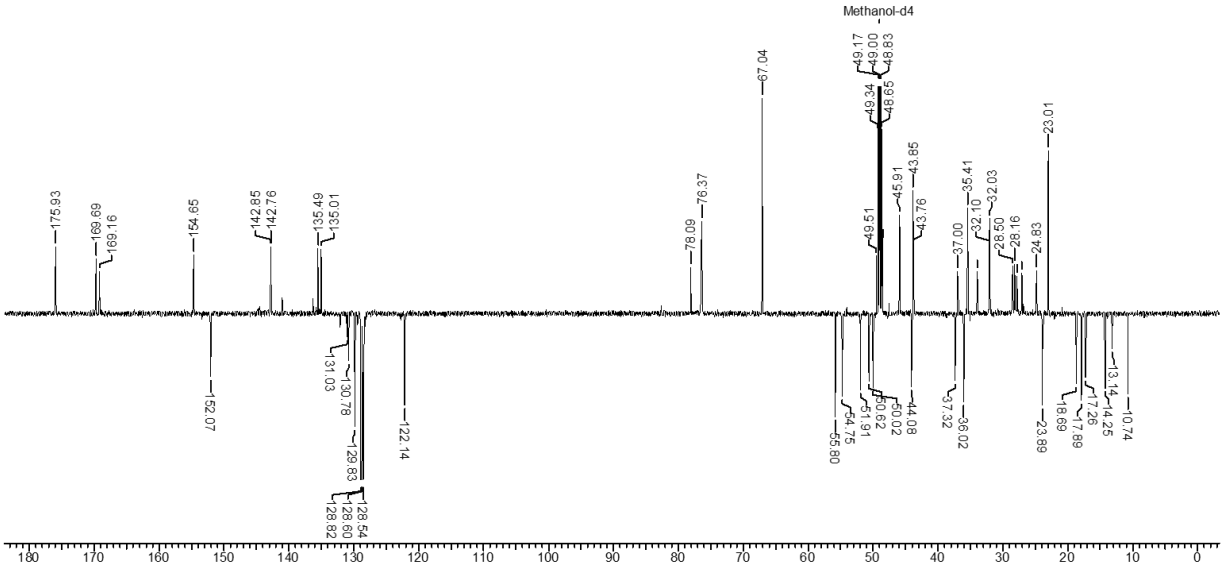


Figure 11.2.13. ^{13}C NMR spectrum of compound **11.2.15**.

11.2.3. Design and Synthesis of linear peptides:

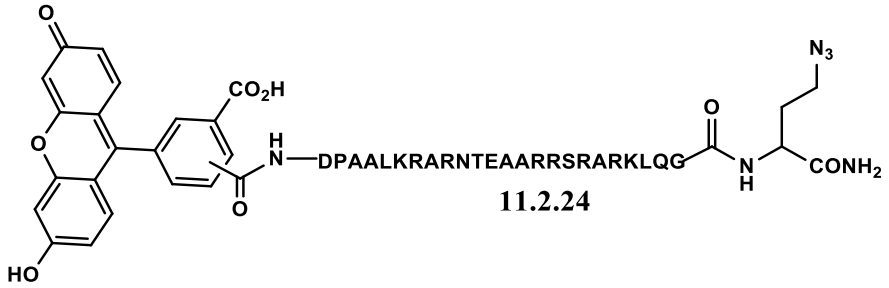
The following six linear peptides were synthesized in context of this study. Since the NHS Fluorescein used is bought in the 5-6 isomer form, all fluorescein containing peptides have a double peak.

11.2.22

Chemical Formula: $C_{123}H_{212}N_{52}O_{35}$
Exact Mass: 2977.64
Molecular Weight: 2979.38

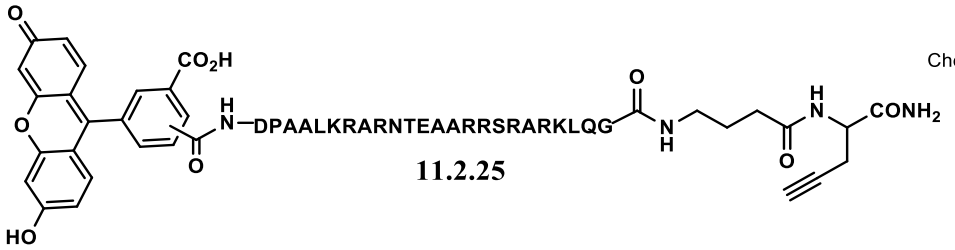
11.2.19

Chemical Formula: $C_{124}H_{211}N_{49}O_{35}$
Exact Mass: 2946.62
Molecular Weight: 2948.36



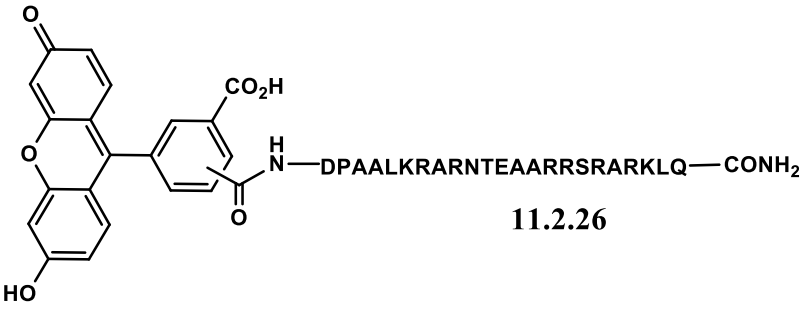
11.2.24

Chemical Formula: $C_{135}H_{215}N_{51}O_{39}$
Exact Mass: 3174.64
Molecular Weight: 3176.52



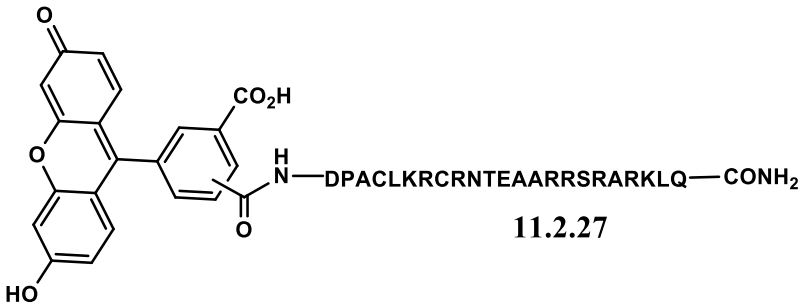
11.2.25

Chemical Formula: $C_{140}H_{221}N_{49}O_{40}$
Exact Mass: 3228.68
Molecular Weight: 3230.61



11.2.26

Chemical Formula: $C_{129}H_{206}N_{46}O_{37}$
Exact Mass: 2991.57
Molecular Weight: 2993.35

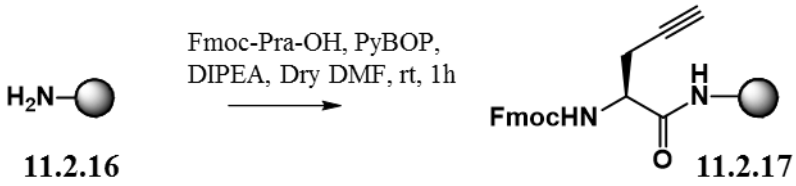


11.2.27

Chemical Formula: $C_{129}H_{206}N_{46}O_{37}S_2$
Exact Mass: 3055.51
Molecular Weight: 3057.47

Manual coupling protocols:

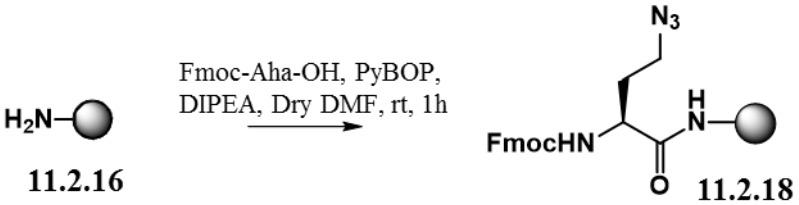
Immobilization of Fmoc-Pra-OH on Rink amide ChemMatrix resin



11.2.16 11.2.17

To a suspension of resin **11.2.16** (50 mg, 0.54 mmol/g, 0.027 mmol) in DMF (dry, 10 mL/g resin), were added Fmoc-Pra-OH (27.2 mg, 0.081 mmol, 4 eq.), PyBOP (42.1 mg, 0.081 mmol, 4 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.). The mixture was shaken at room temperature for 1 h. After the reaction, the resin was washed with DMF, ACN and MeOH. The loading was 100%. The resin was then capped with acetic anhydride (0.016 mL, 0.162 mmol, 8 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.) in dry DMF (2 mL) twice for 30 min.

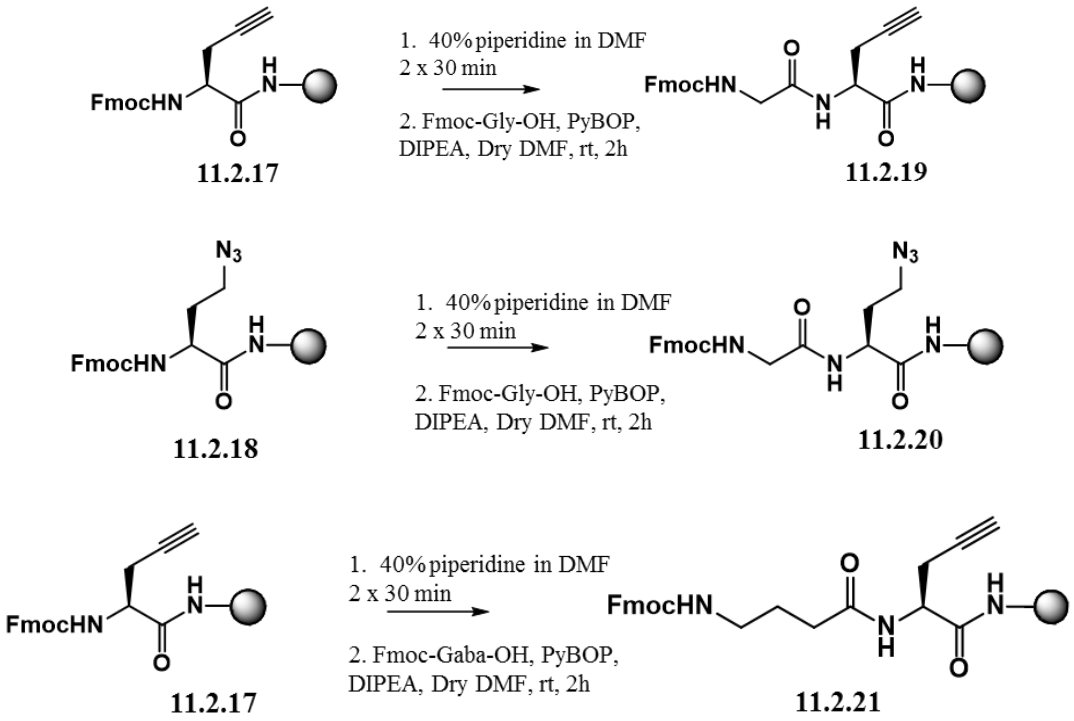
Immobilization of Fmoc-Aha-OH on Rink amide ChemMatrix resin



11.2.16 11.2.18

To a suspension of resin **11.2.16** (50 mg, 0.54 mmol/g, 0.027 mmol) in DMF (dry, 10 mL/g resin), were added Fmoc-Aha-OH (29.7 mg, 0.081 mmol, 4 eq.), PyBOP (42.1 mg, 0.081 mmol, 4 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.). The mixture was shaken at room temperature for 1 h. After the reaction, the resin was washed with DMF, ACN and MeOH and dried in vacuum for 1h. The loading was calculated to be 0.412 mmol/g, indicating a coupling yield of 92%. The resin was then capped with acetic anhydride (0.016 mL, 0.162 mmol) and DIPEA (0.028 mL, 0.162 mmol) in dry DMF (2 mL) 2 times for 30 min.

Fmoc deprotection and coupling of Fmoc-Gly-OH and Fmoc-GABA-OH



Fmoc deprotection of compounds **11.2.17** & **11.2.20**. After an initial DMF washing step, resin **11.2.17** (0.022 mmol) was treated twice successively for 30 min with a piperidine solution in DMF (40 % v/v, 2 mL) at ambient temperature, applying intermediate filtration under reduced pressure and washing with 3 x DMF, 3 x ACN, 3 x DCM and 3 x DMF.

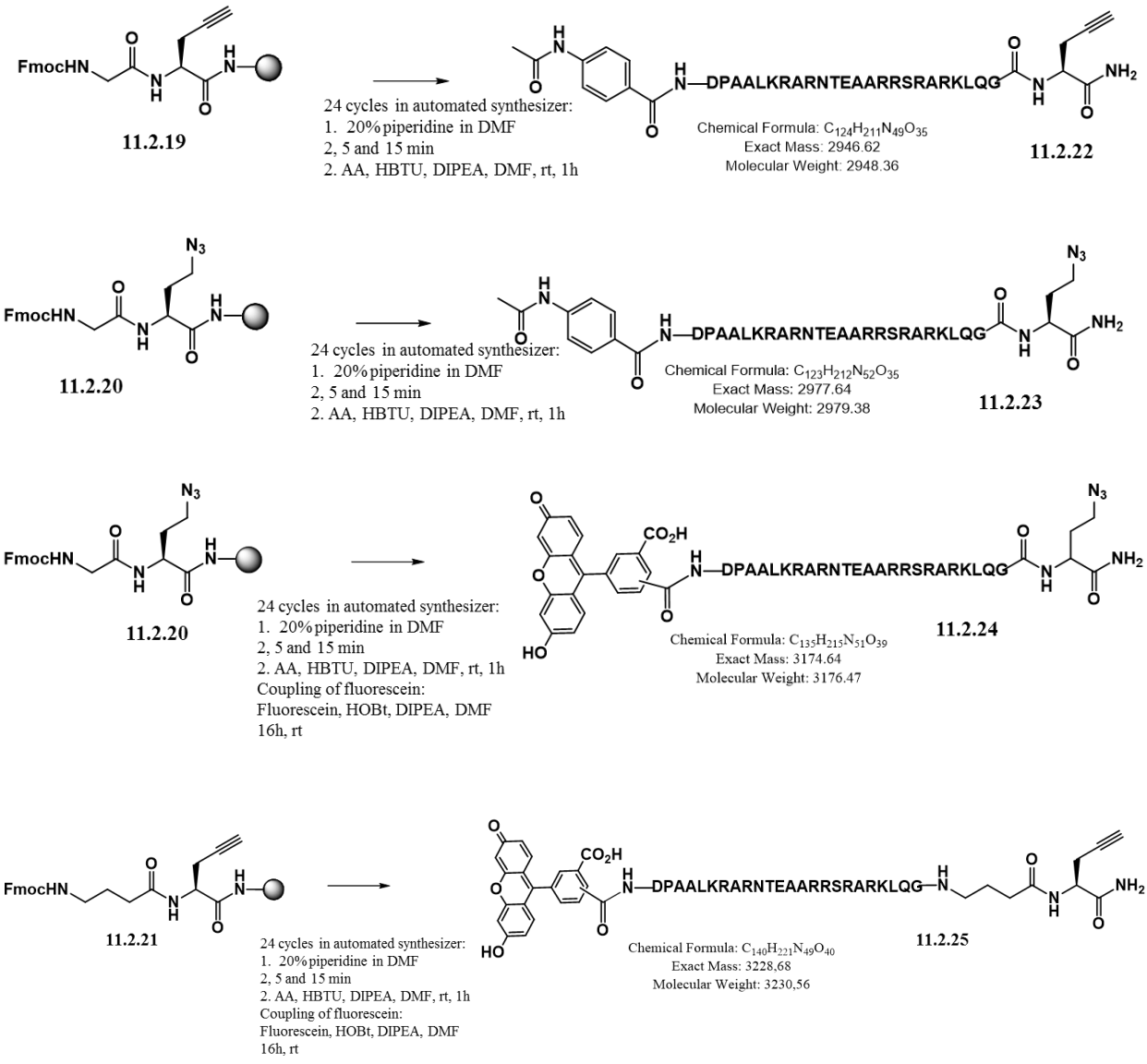
Coupling of Fmoc-Gly-OH to **11.2.17**. Fmoc-Gly-OH (19.5 mg, 0.065 mmol, 0.5 M), PyBOP (34.15 mg, 0.065 mmol, 0.5 M) and DIPEA (0.012 mL, 0.065 mmol, 2 M) were added to a suspension of resin in dry DMF (2 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

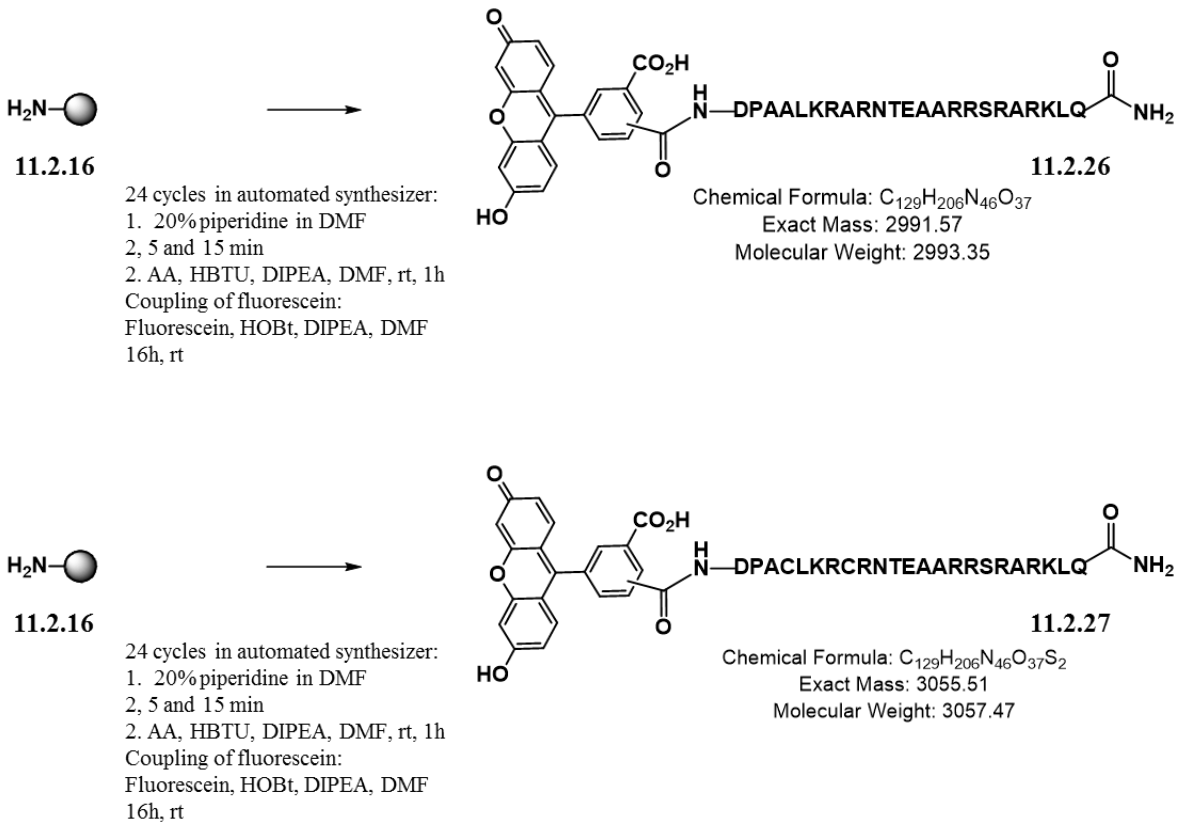
Coupling of Fmoc-Gly-OH to **11.2.20**. Fmoc-Gly-OH (22.3 mg, 0.075 mmol, 0.5 M), PyBOP (39 mg, 0.075 mmol, 0.5 M) and DIPEA (0.013 mL, 0.075 mmol, 2 M) were added to a suspension of resin in dry DMF (2 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

Coupling of Fmoc-GABA-OH. To Fmoc deprotected resin (0.025 mmol), Fmoc-Gaba-OH (35.1 mg, 0.108 mmol, 0.5 M), PyBOP (56.2 mg, 0.108 mmol, 0.5 M) and DIPEA (0.019 mL, 0.108 mmol, 2 M) were added to a suspension of resin in dry DMF (2 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

Coupling of Fluorescein: 3 eq. of NHS-fluorescein predissolved in 0.5 mL DMF. 6 eq. of DIPEA and 3 eq. of HOBt predissolved in 0.5 mL DMF and are added successively to the resin. After shaking for 15 h, the reaction mixture was filtered off and the resin washed with 3x DMF, 3x MeOH, 3x DMF and 3x Et_2O.

Automated synthesis of linear peptides:





Automated solid phase peptide synthesis was carried out according to protocols described in sections 10.2.

The peptides were simultaneously cleaved from the resin and deprotected with a cocktail of TFA/TIS/water = 95: 2.5: 2.5 for 4h (4 mL/50 mg). After cleavage the resin was washed with neat TFA and was precipitated in cold ether (-20°C) by centrifugation at 0-4°C. The peptides were analyzed by RP-HPLC/LC-MS and/or MALDI.

11.2.4. Analysis of linear peptides

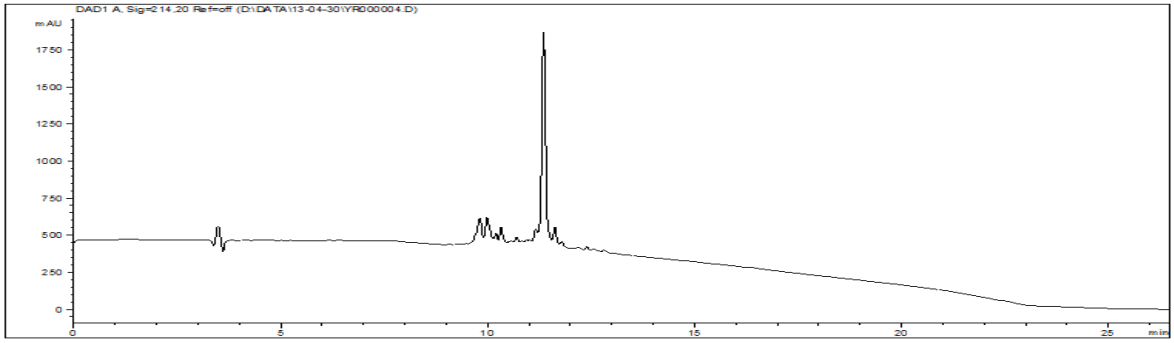


Figure 11.2.14: RP-HPLC Chromatogram of crude compound **11.2.19** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

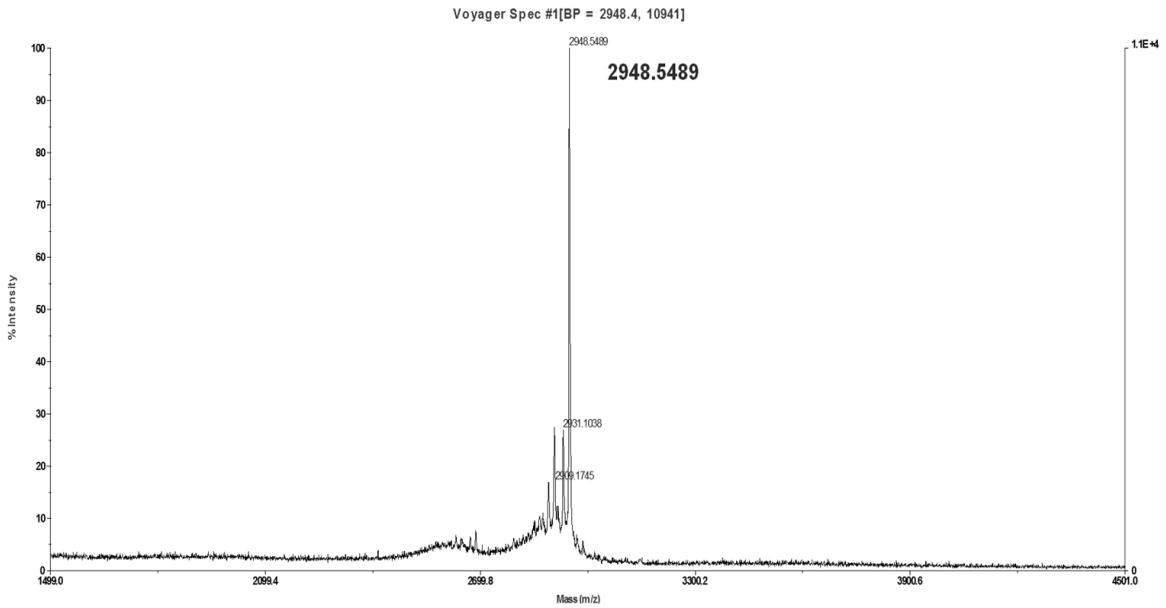


Figure 11.2.15: MALDI-TOF Spectrum of crude compound **11.2.19**. Calculated m/z (100%) = 2947.63; m/z (73.9%) = 2948.63. Found: 2948.5489 = M + H^+.

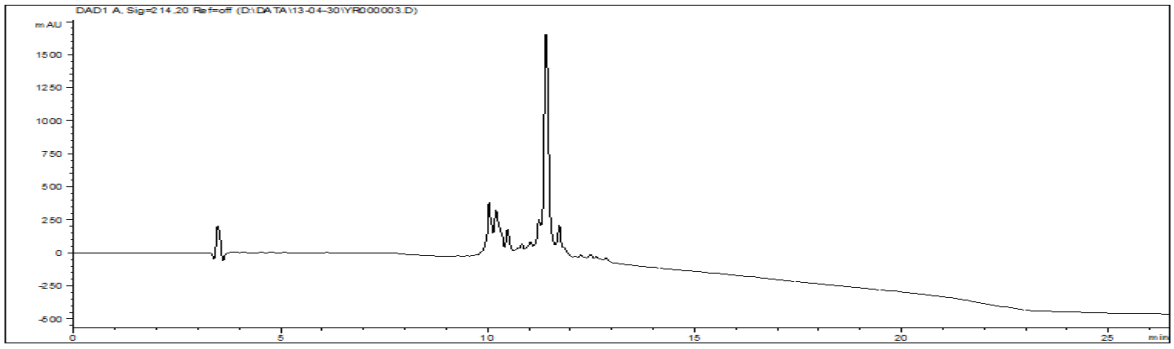


Figure 11.2.16: RP-HPLC Chromatogram of crude compound **11.2.22** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

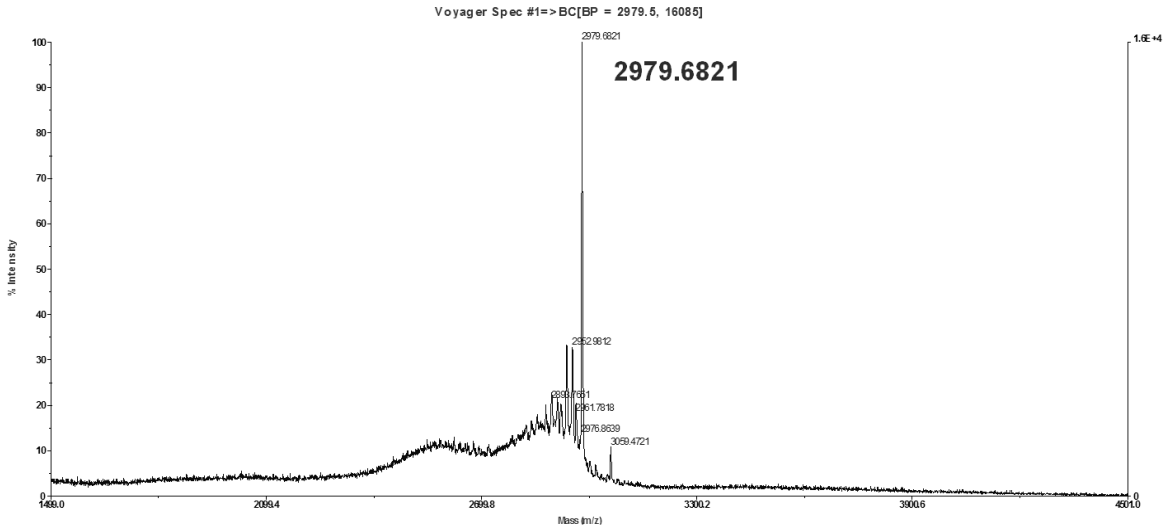


Figure 11.2.17: MALDI-TOF Spectrum of crude compound **11.2.22**. Calculated m/z (100%) = 2978.64; m/z (65.1%) = 2979.65. Found: 2979.6821 = M + H^+.

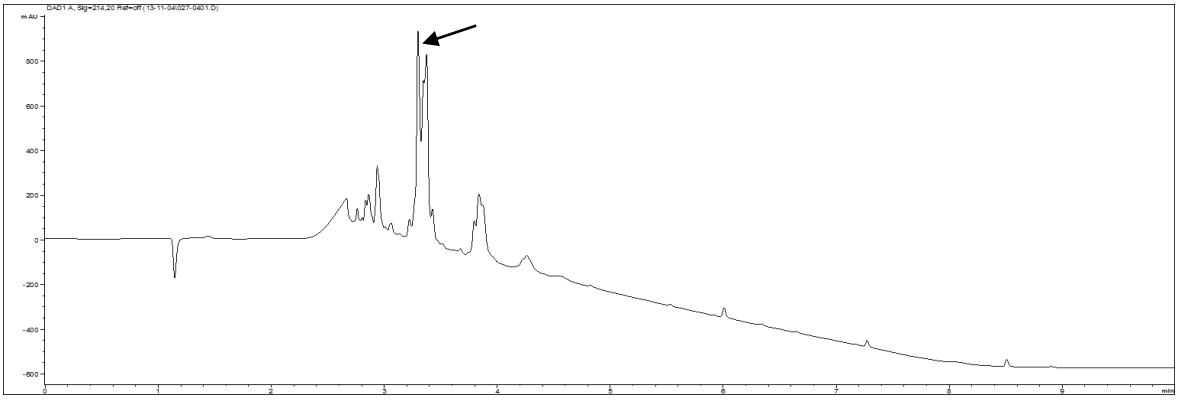


Figure 11.2.18: RP-HPLC of compound **11.2.24** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (Double peak due to the isomers of fluorescein).

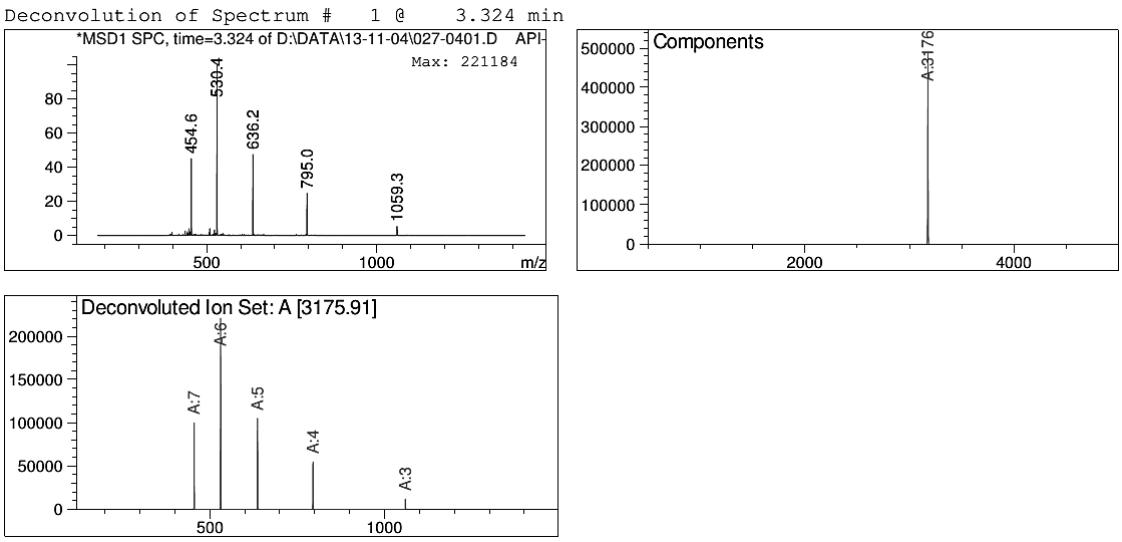


Figure 11.2.19: ESI-MS from LC-MS at t = 3.324 min for crude compound **11.2.24**. E.M calcd. for $C_{135}H_{215}N_{51}O_{39}$ = 3174.64 and deconvoluted mass found 3175.91.

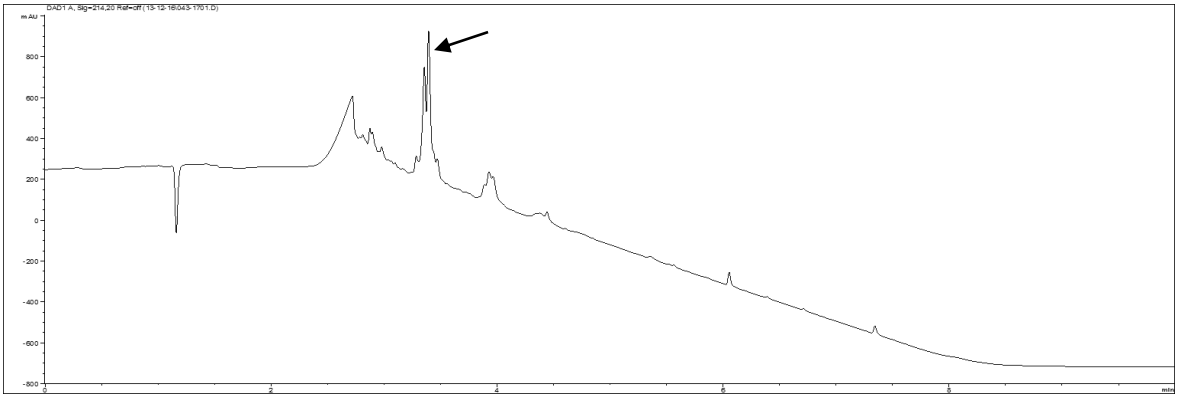


Figure 11.2.20: RP-HPLC of compound **11.2.25** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (Double peak due to the isomers of fluorescein).

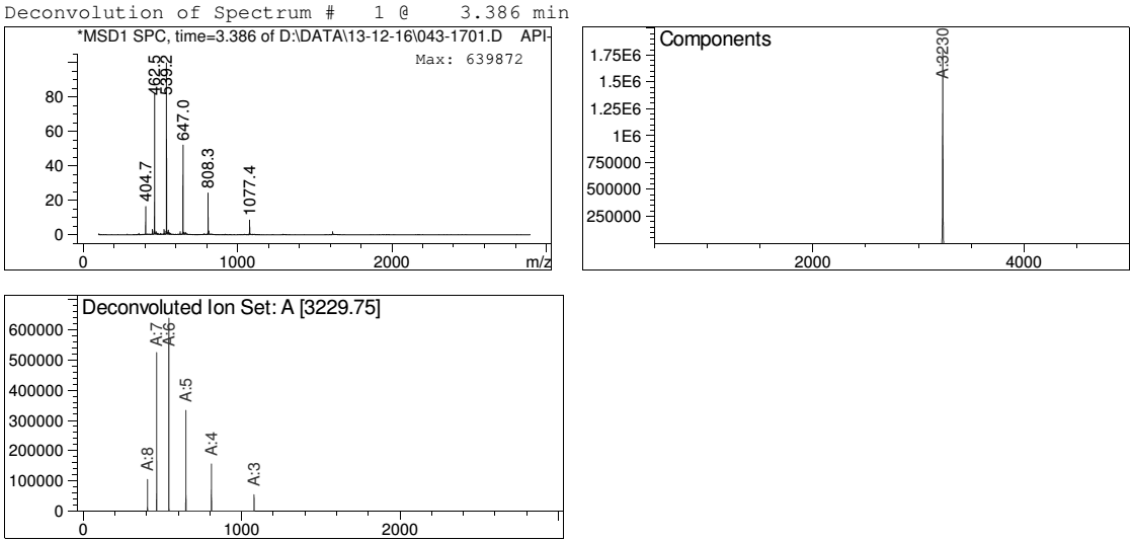


Figure 11.2.21: ESI-MS from LC-MS at t = 3.324 min for crude compound **11.2.25**. E.M calcd. for $C_{140}H_{221}N_{49}O_{40}$ = 3228.68 and deconvoluted mass found 3229.75.

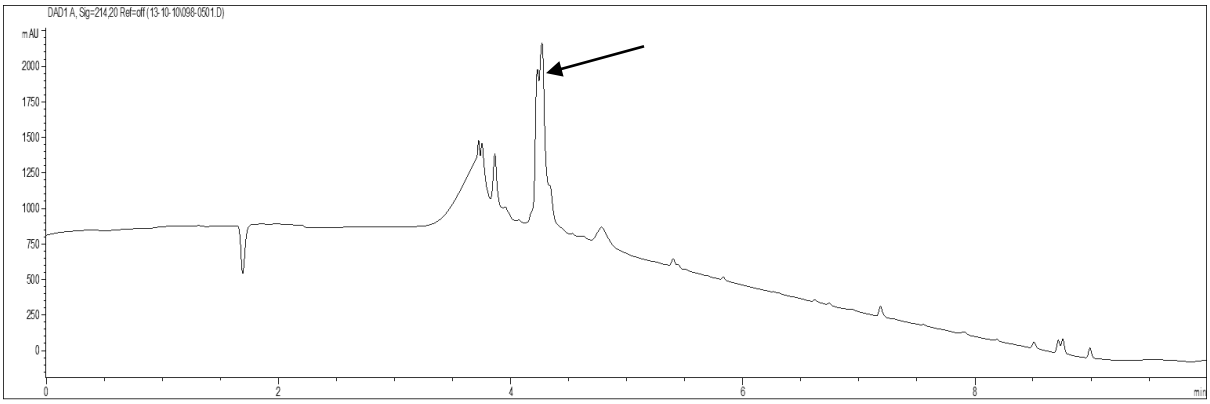


Figure 11.2.22: RP-HPLC of compound **11.2.26** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (Double peak due to the isomers of fluorescein).

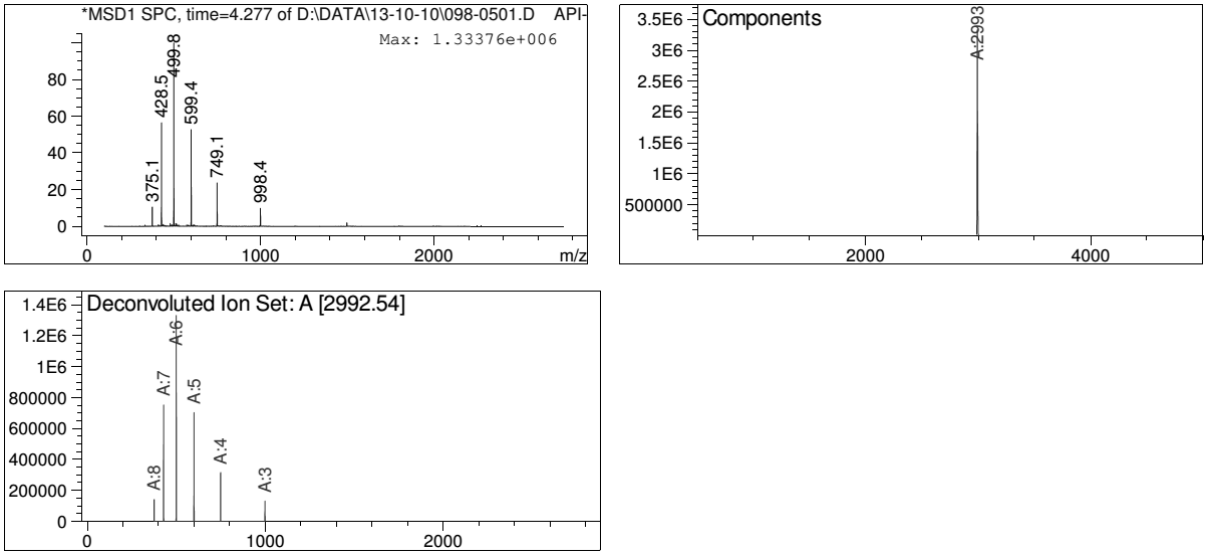


Figure **11.2.23**: ESI-MS from LC-MS at t = 4.277 min for crude compound **11.2.26**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.54.

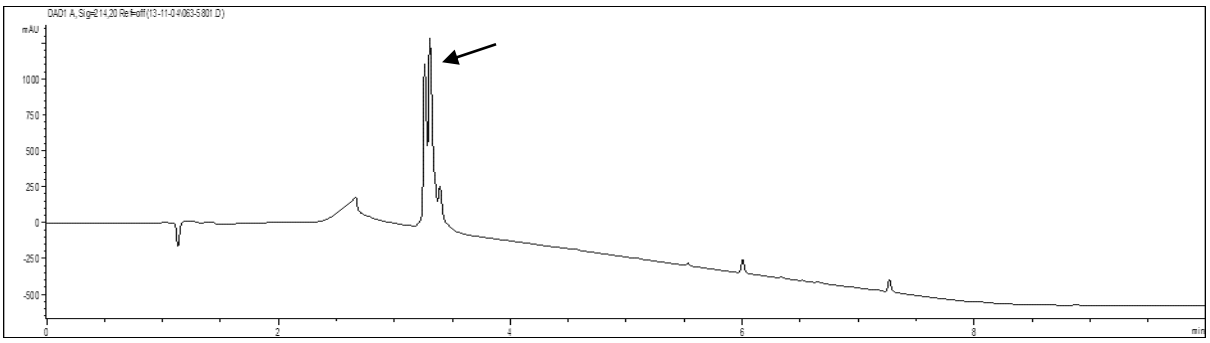


Figure 11.2.24: RP-HPLC trace of HPLC purified compound **11.2.26** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (Double peak due to the isomers of fluorescein).

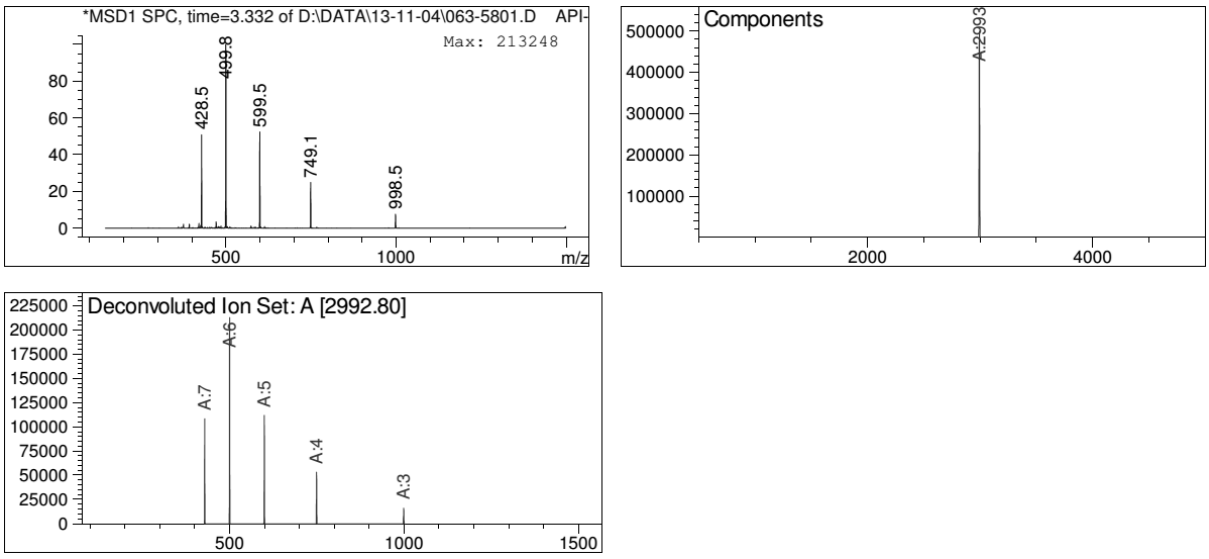


Figure 11.2.25: ESI-MS from LC-MS at t = 3.332 min. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.80.

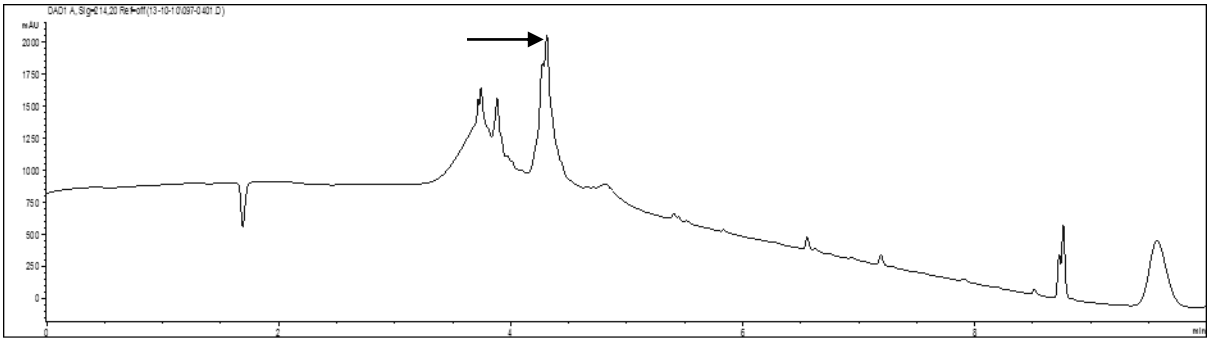


Figure 11.2.26: RP-HPLC trace of HPLC crude compound **11.2.27** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (Double peak due to the isomers of fluorescein).

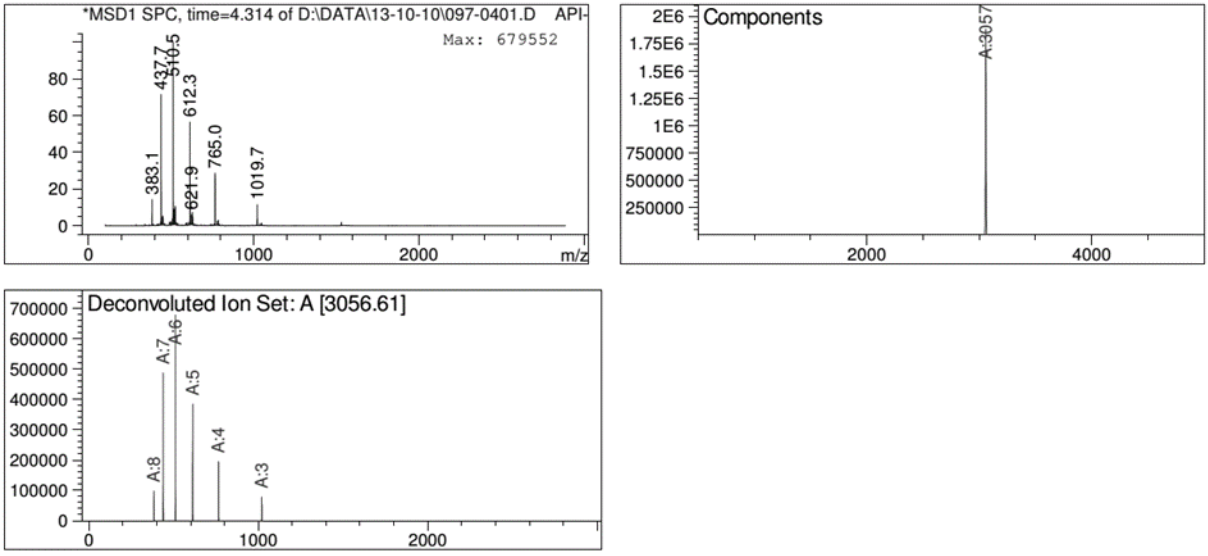
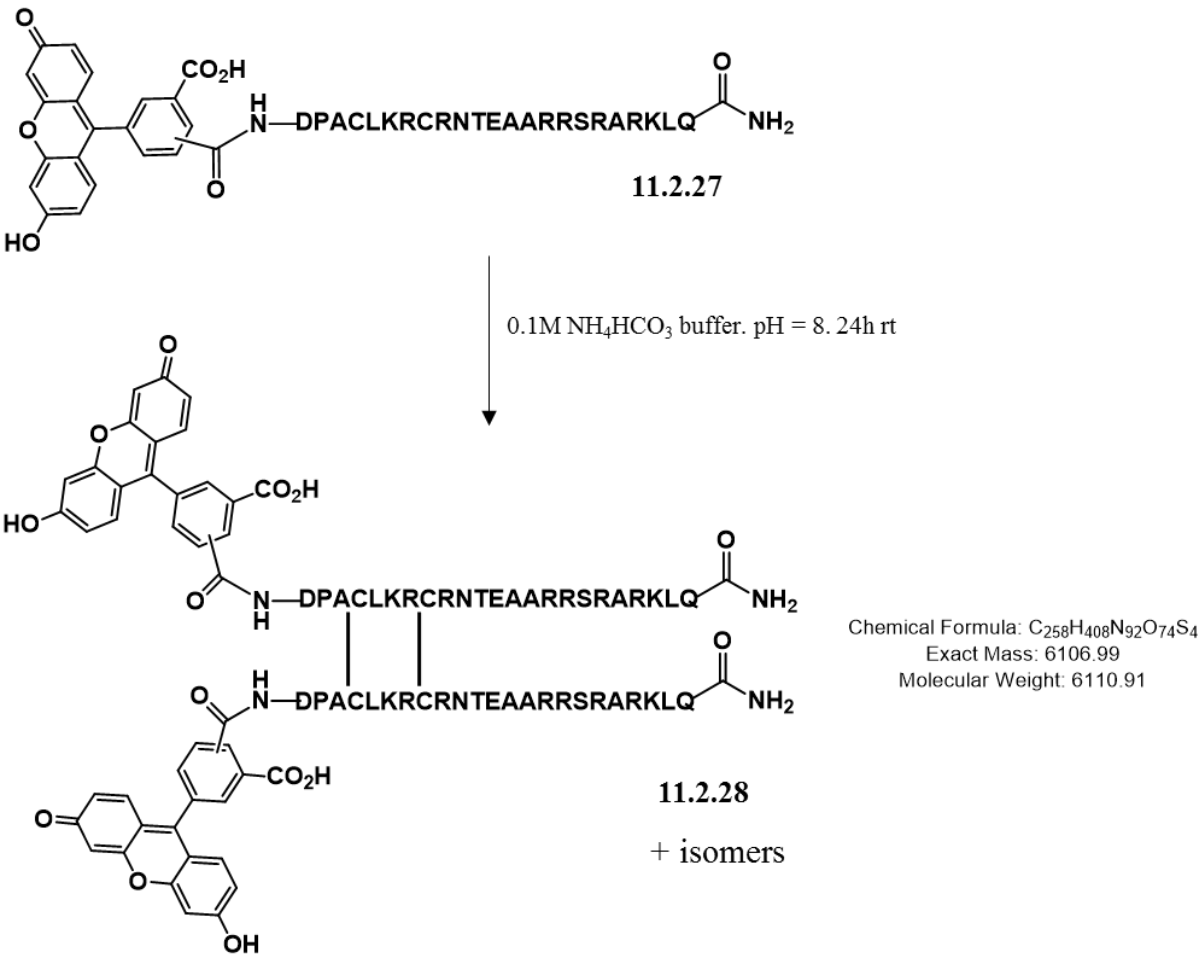


Figure 11.2.27: ESI-MS from LC-MS at t = 4.314 min for crude compound **11.2.27**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}S_2$ = 3055.51 and deconvoluted mass found 3056.61.

Synthesis of compound 11.2.28



Peptide **11.2.27** (5 mg, 1.63 µmol) was dissolved in 0.1 M NH_4HCO_3 buffer (1 mL) and stirred vigorously for 24h. The reaction was then quenched with TFA. A mixture of peptides containing 2 peptide strands but with single or double disulfide bonds was isolated using RP-HPLC. No higher mass products were observed.

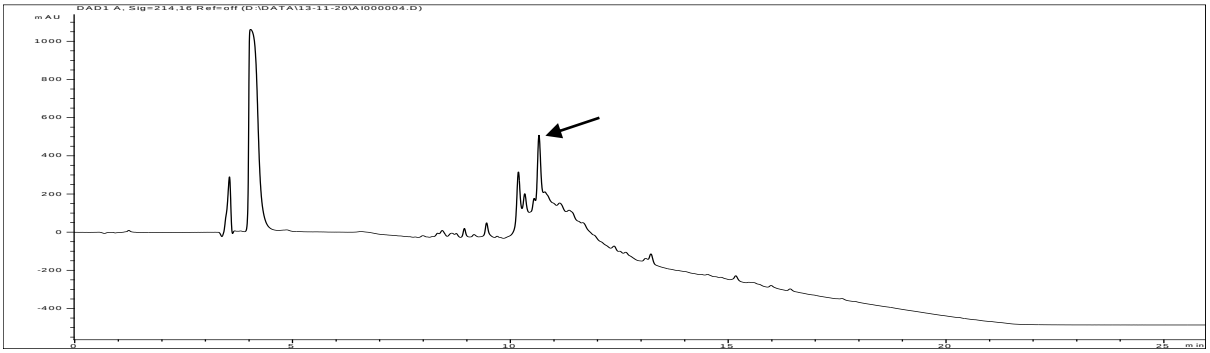


Figure 11.2.28: RP-HPLC Chromatogram of reaction mixture showing product **11.2.8** at t = 10.65 min (C18, 100Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

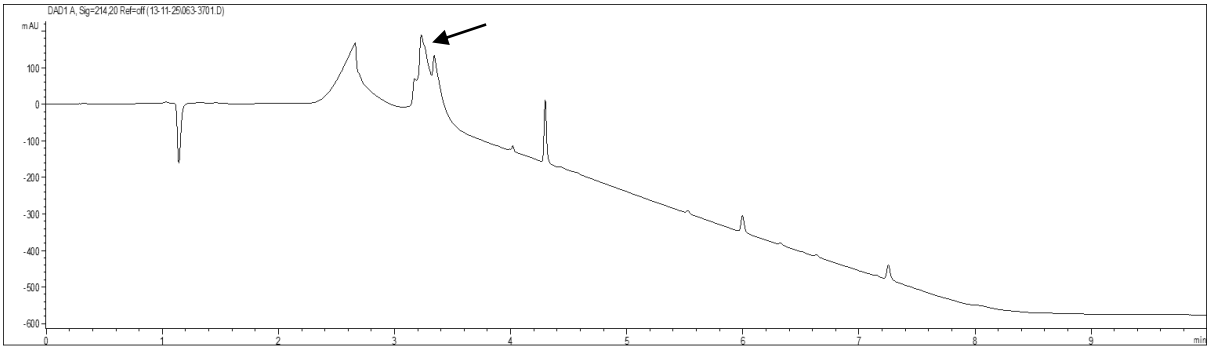


Figure 11.2.29: RP-HPLC trace of HPLC purified compound **11.2.28** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

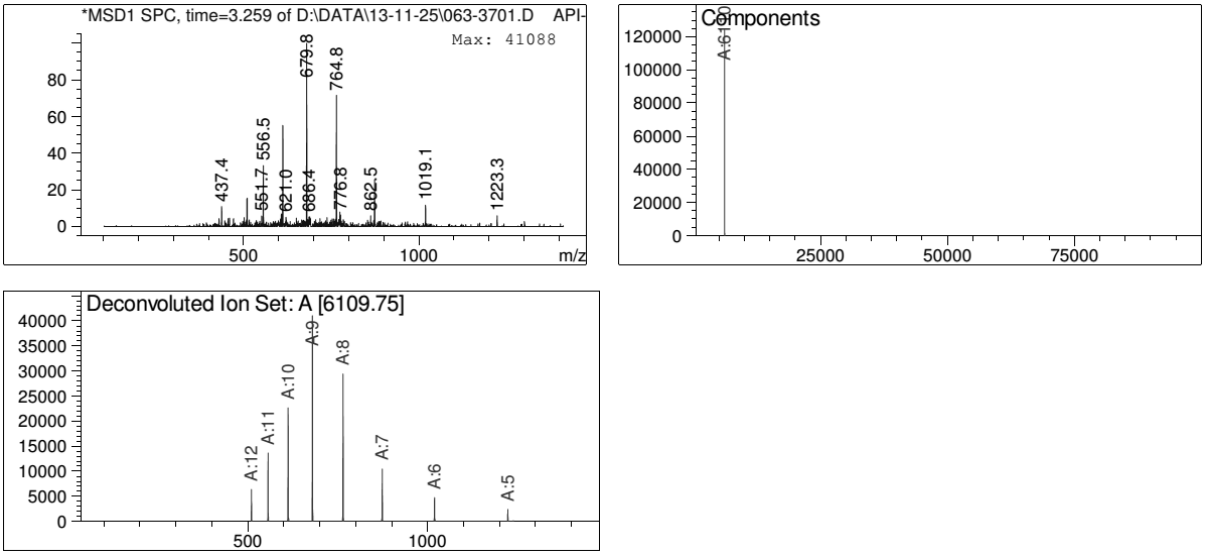


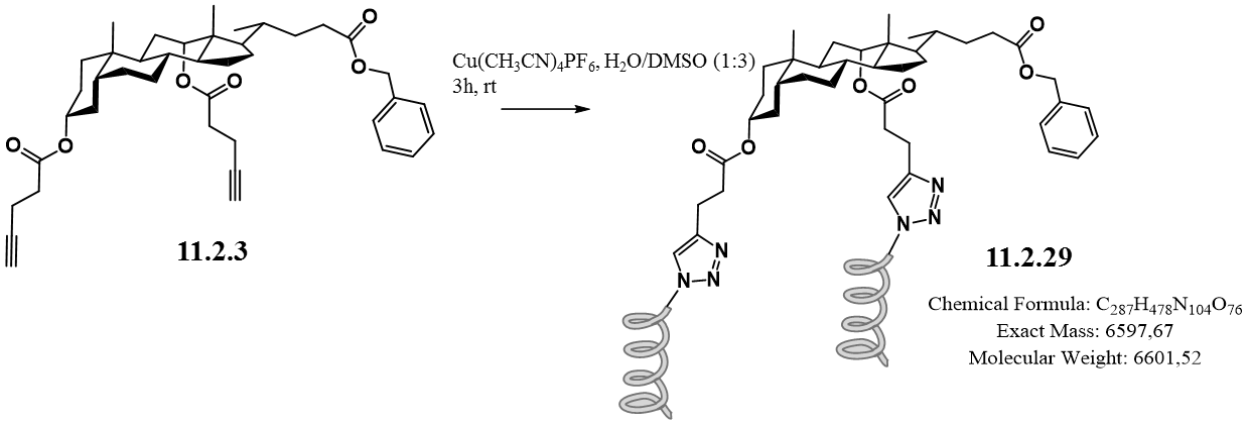
Figure 11.2.30: ESI-MS from LC-MS at t = 3.259 min for purified compound **11.2.28**. E.M calcd. for $C_{258}H_{408}N_{92}O_{74}S_4$ = 6106.99 and $C_{258}H_{410}N_{92}O_{74}S_4$ = 6109.00 and deconvoluted mass found 6109.75.

11.2.5 CuAAC Conjugation I: peptides & steroids

Scaffold-Peptide Conjugation Protocol via CuAAC

COMPOUNDS FOR DNA BINDING

Pentynoic acid (11.2.29): Scaffold **11.2.3** (3.2 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.23** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.63 mg, 12 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



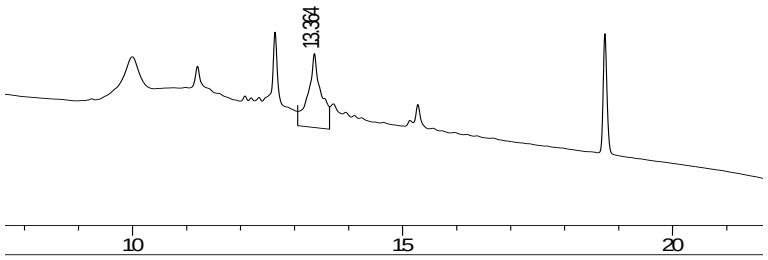


Figure 11.2.31: RP-HPLC Chromatogram of crude compound **11.2.29** eluting at 13.3 min (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

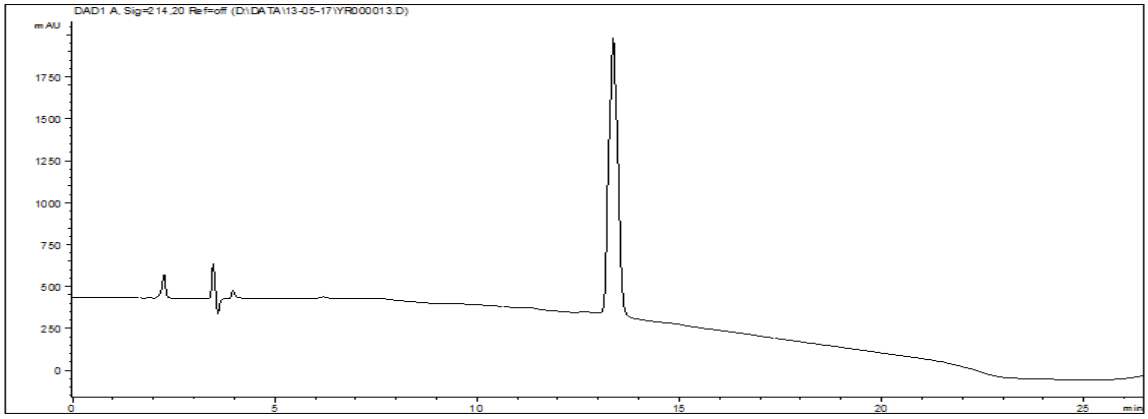


Figure 11.2.32: RP-HPLC Chromatogram of pure compound **11.2.29** (C4, 300Å column using a gradient from 0 to 100 % CH3CN in 15 minutes).

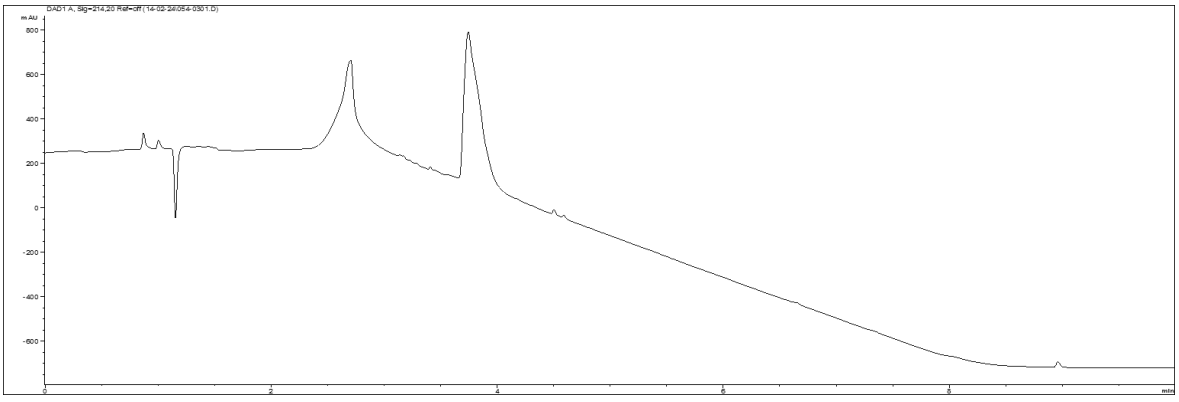


Figure 11.2.33: RP-HPLC of compound **11.2.29** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

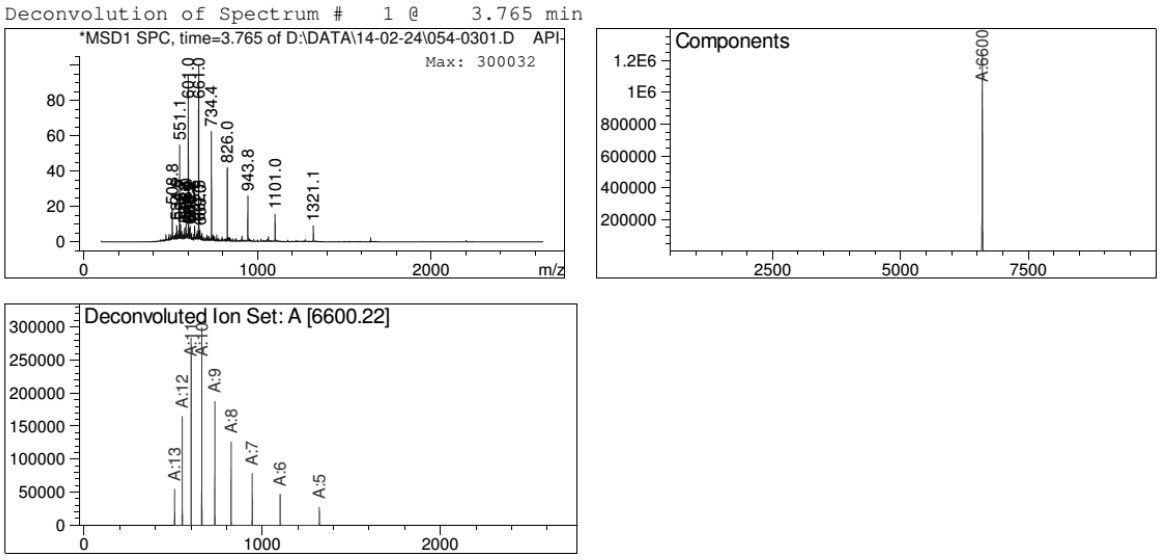
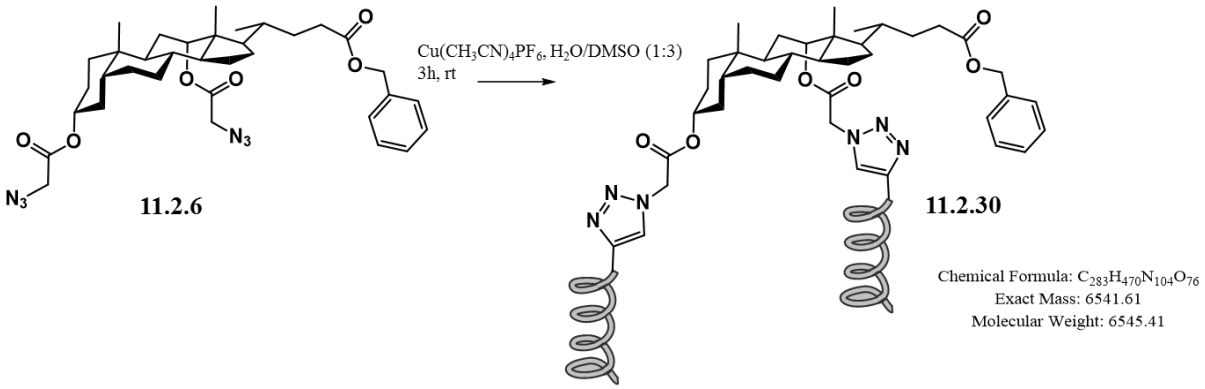


Figure 11.2.34: ESI-MS spectrum from LC-MS at t = 3.765 min of pure compound **11.2.23**. EM calcd. for $C_{287}H_{478}N_{104}O_{76}$ = 6597.67 and deconvoluted mass found 6600.2.

Gly (11.2.30): Scaffold **11.2.6** (3.2 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.22** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.68 mg, 13 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



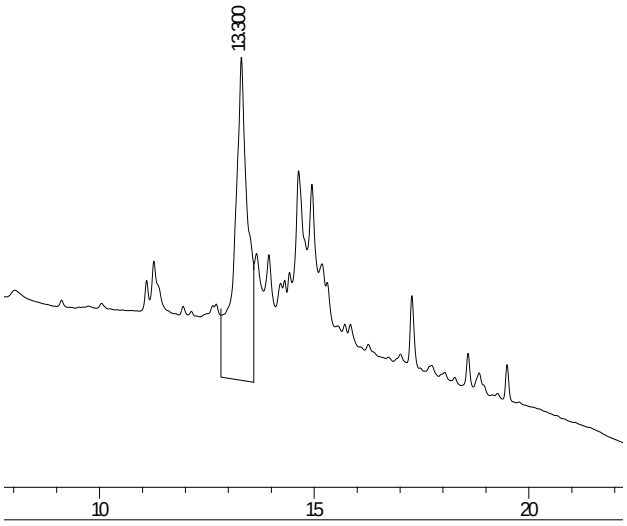


Figure 11.2.35: RP-HPLC Chromatogram of crude compound **11.2.30** eluting at 13.3 min (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

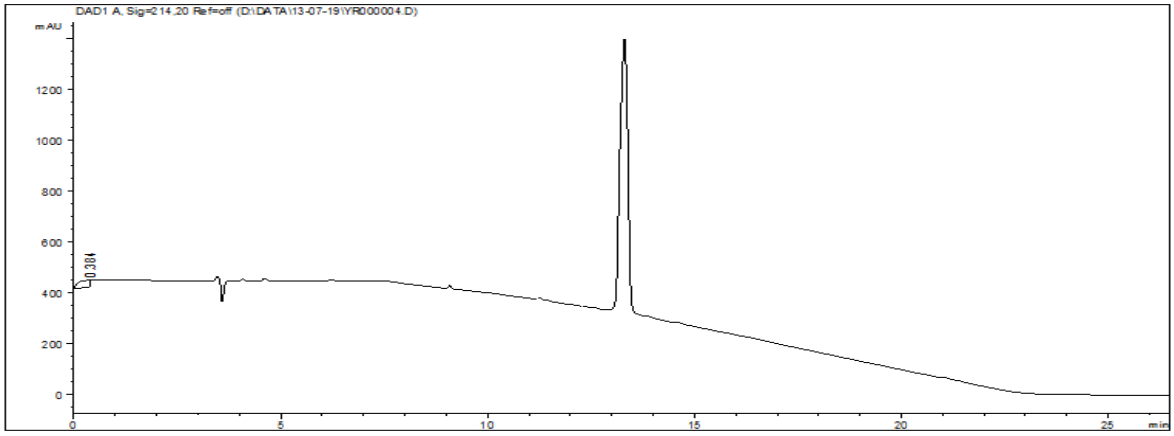


Figure 11.2.36: RP-HPLC Chromatogram of pure compound **11.2.30** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

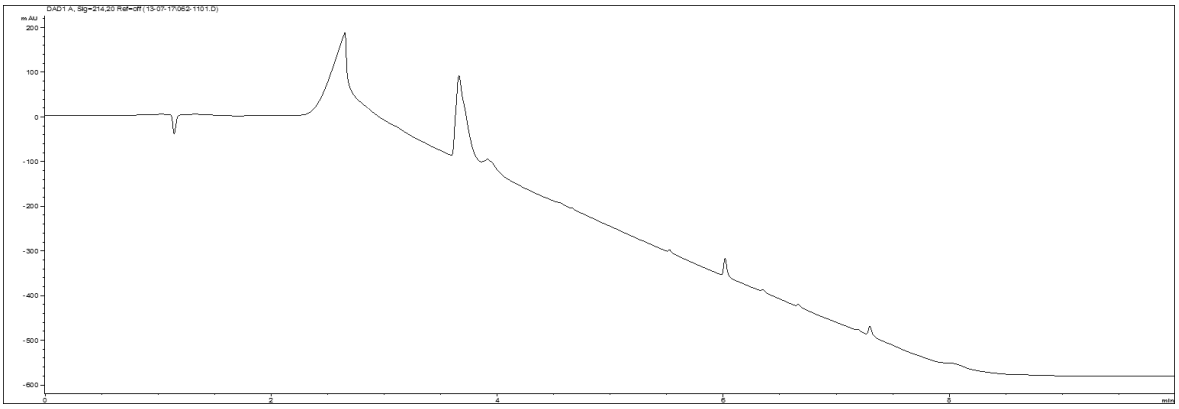


Figure 11.2.37: RP-HPLC trace of compound **11.2.30** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

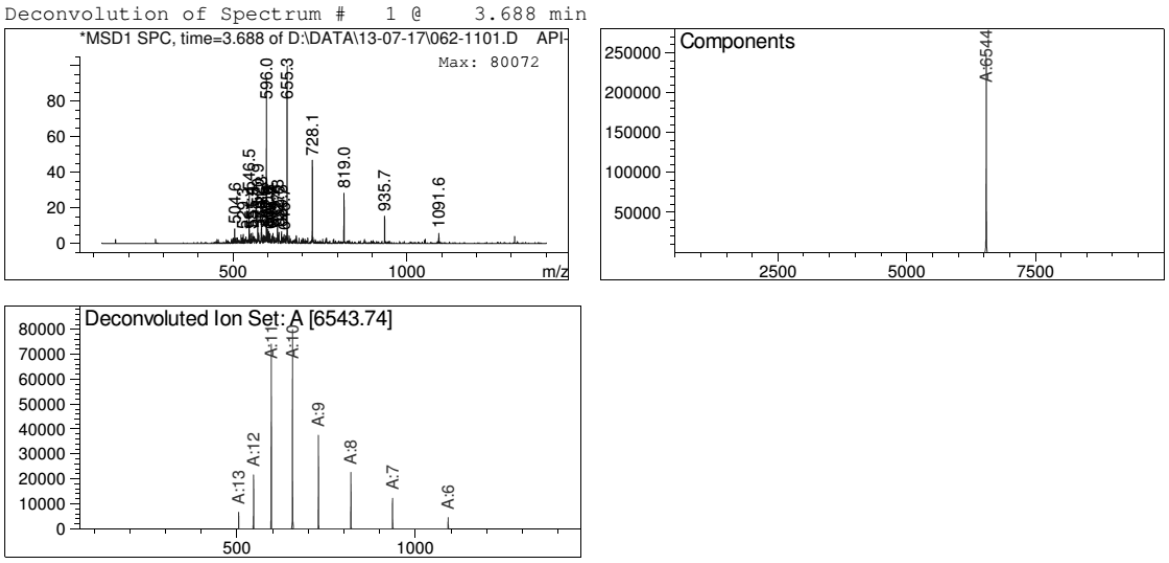


Figure 11.2.38: ESI-MS spectrum from LC-MS at t = 3.688 min of compound **11.2.30**. EM cacld. for $C_{283}H_{470}N_{104}O_{76}$ = 6541.61 and deconvoluted mass found 6543.74.

Azidobenzoic acid (11.2.31): Scaffold **11.2.9** (4 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.22** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.68 mg, 13 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.

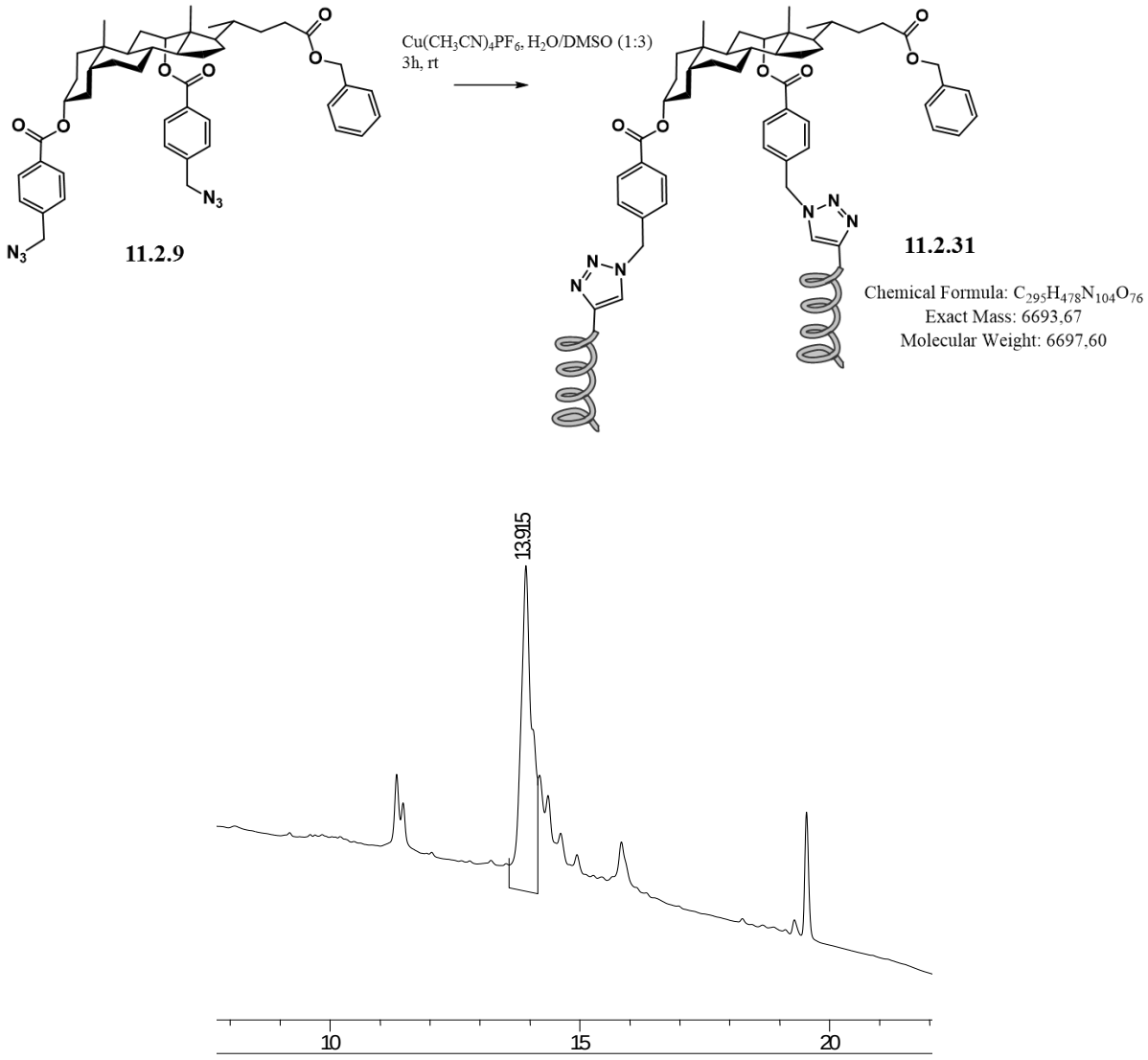


Figure 11.2.39: RP-HPLC Chromatogram of crude compound **11.2.31** eluting at 13.9 min (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

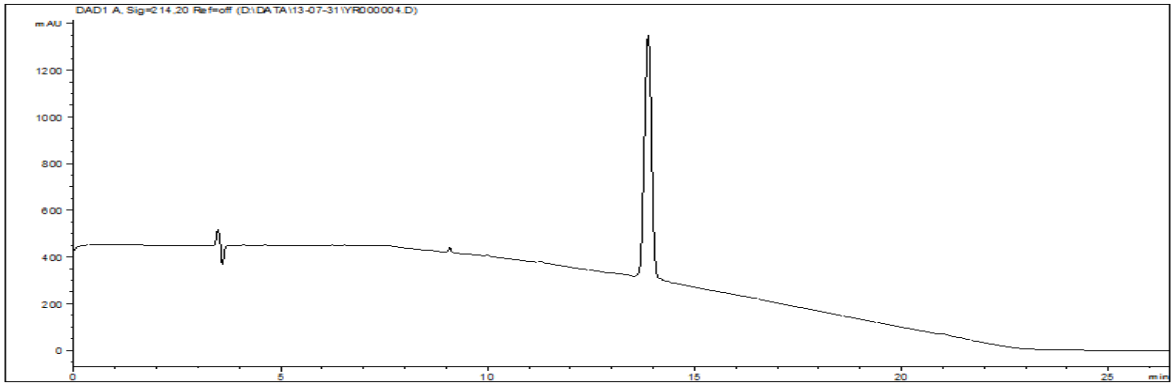


Figure 11.2.40: RP-HPLC Chromatogram of pure compound **11.2.31** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

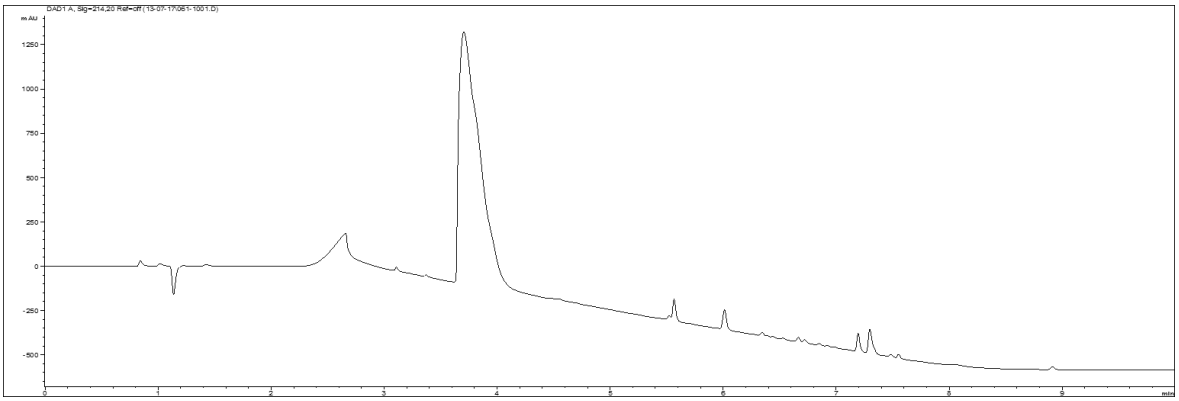


Figure 11.2.41: RP-HPLC trace of compound **11.2.31** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

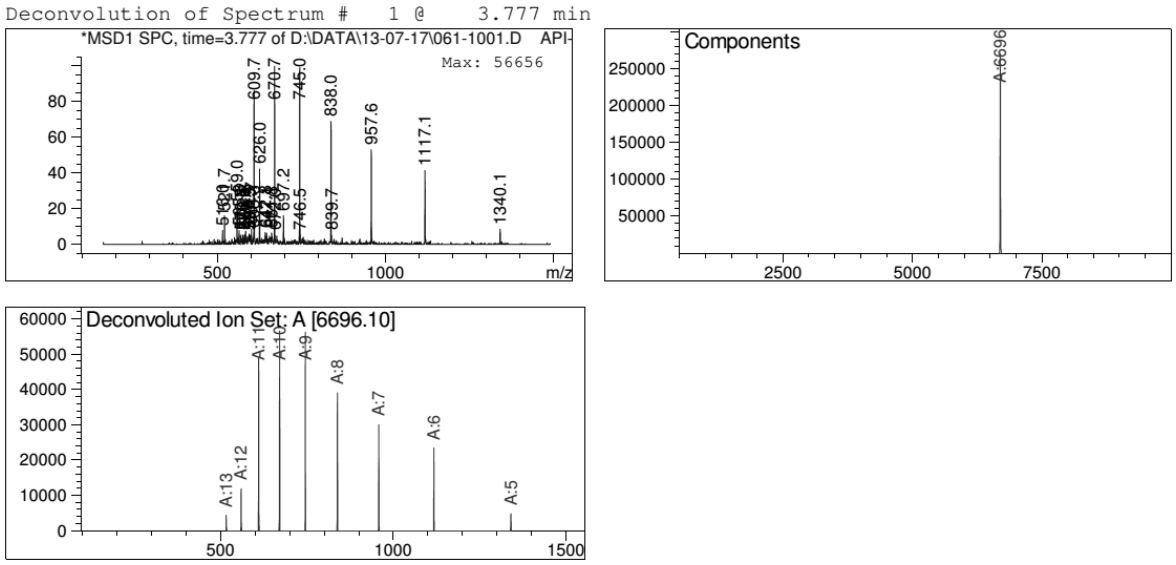


Figure 11.2.42: ESI-MS from LC-MS at t = 3.777 min of compound **11.2.31**. EM calcd. for $C_{295}H_{478}N_{104}O_{76}$ = 6693.97 and deconvoluted mass found 6696.10.

PATA linker (11.2.32): Scaffold **11.2.15** (4 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.22** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.63 mg, 12 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and LC-MS.

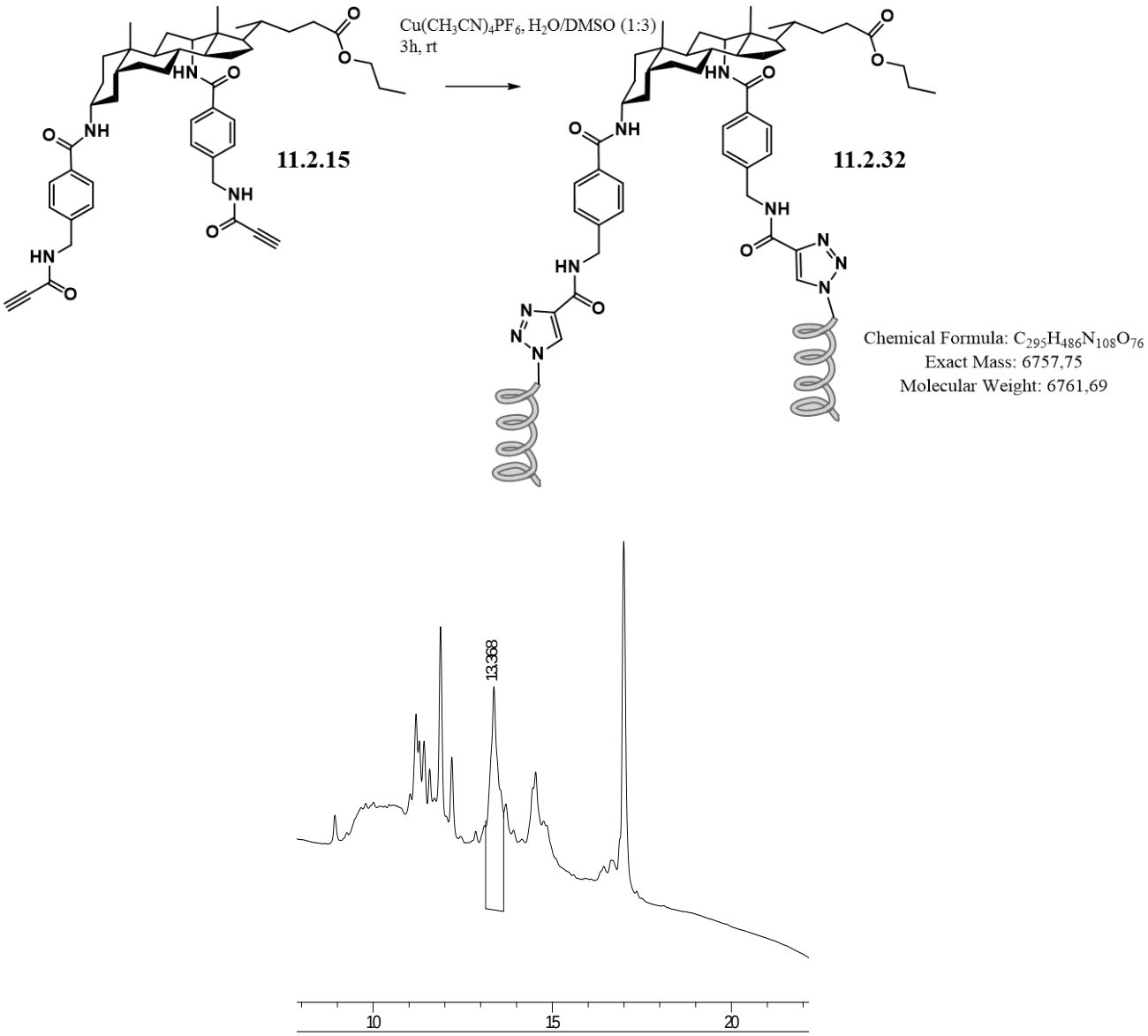


Figure 11.2.43: RP-HPLC Chromatogram of crude compound **11.2.32** eluting at 13.368 min (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

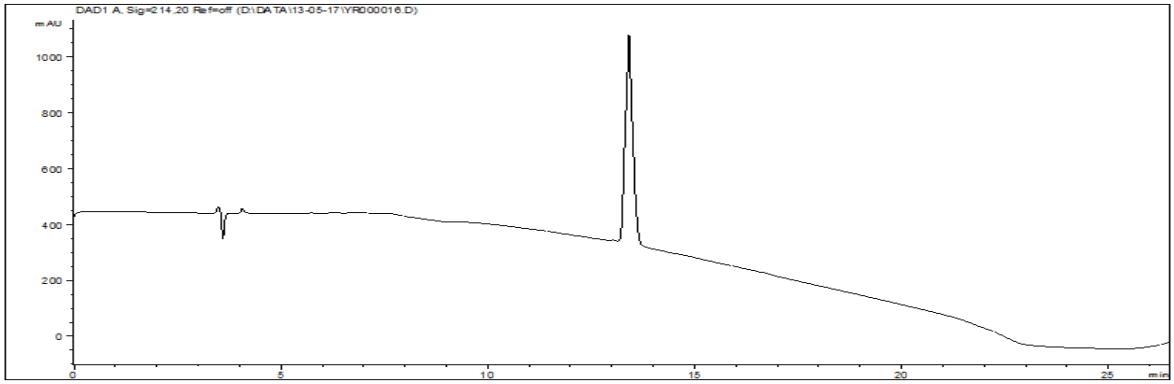


Figure 11.2.44: RP-HPLC Chromatogram of pure compound **11.2.32** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

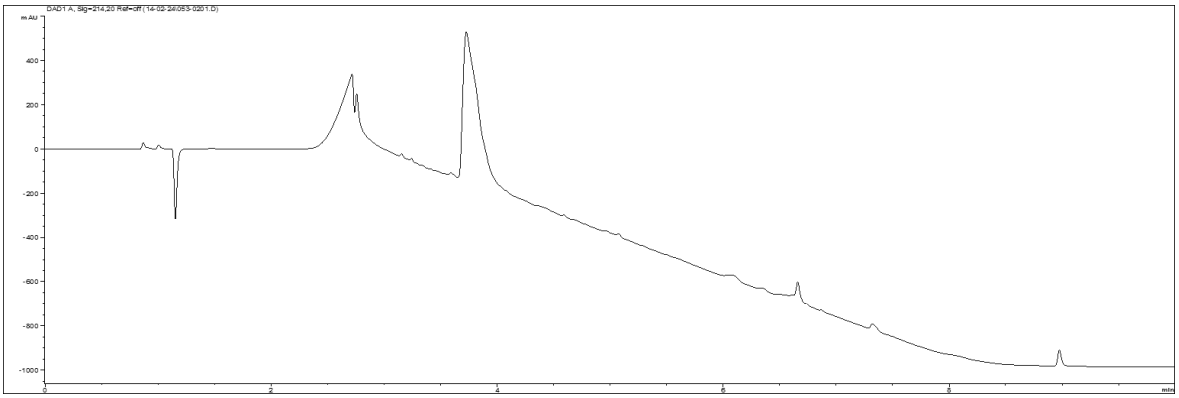


Figure 11.2.45: RP-HPLC of compound **11.2.32** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

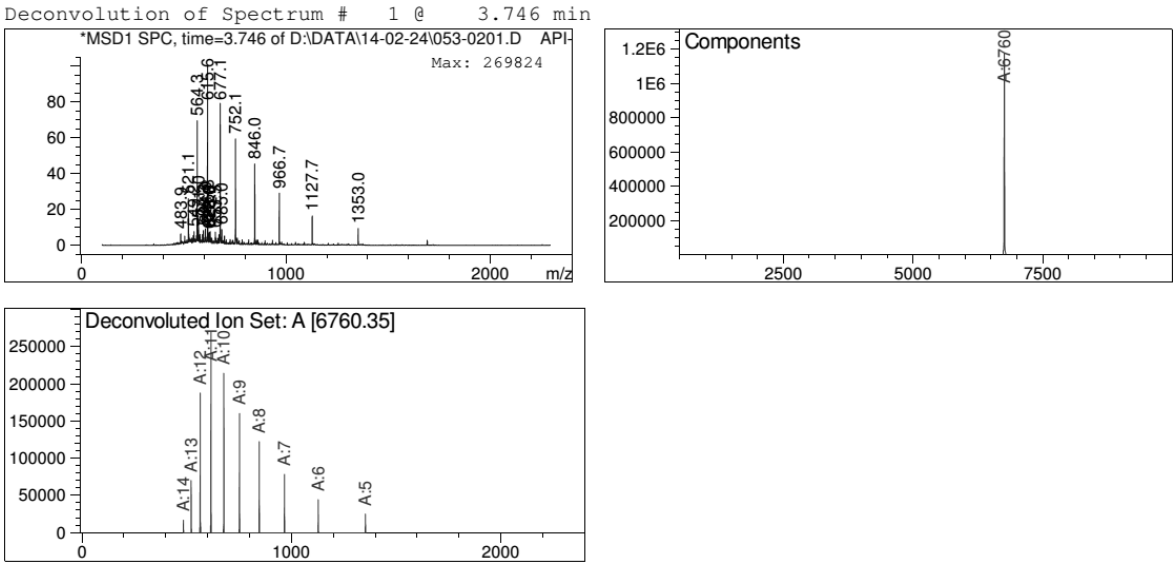
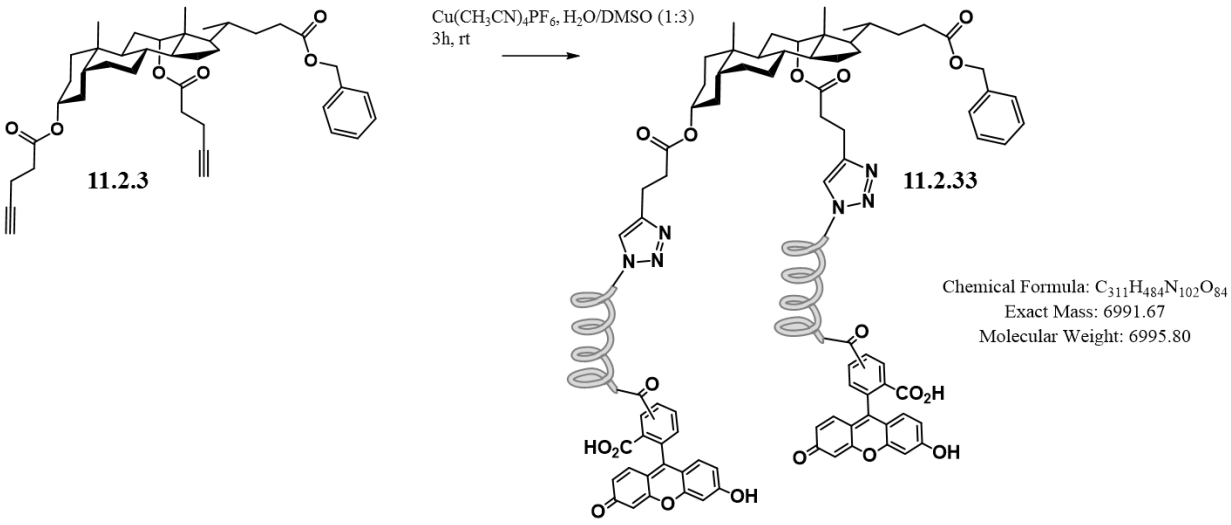


Figure 11.2.46: ESI-MS spectrum from LC-MS at t = 3.746 min of compound **11.2.26**. EM calcd. for $C_{295}H_{486}N_{108}O_{76}$ = 6757.75 and deconvoluted mass found 6760.35.

COMPOUNDS FOR CELL PENETRATION

Synthesis of 11.2.33: Scaffold **11.2.3** (0.27 mg, 0.4 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.24** (4 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (2.34 mg, 6.3 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



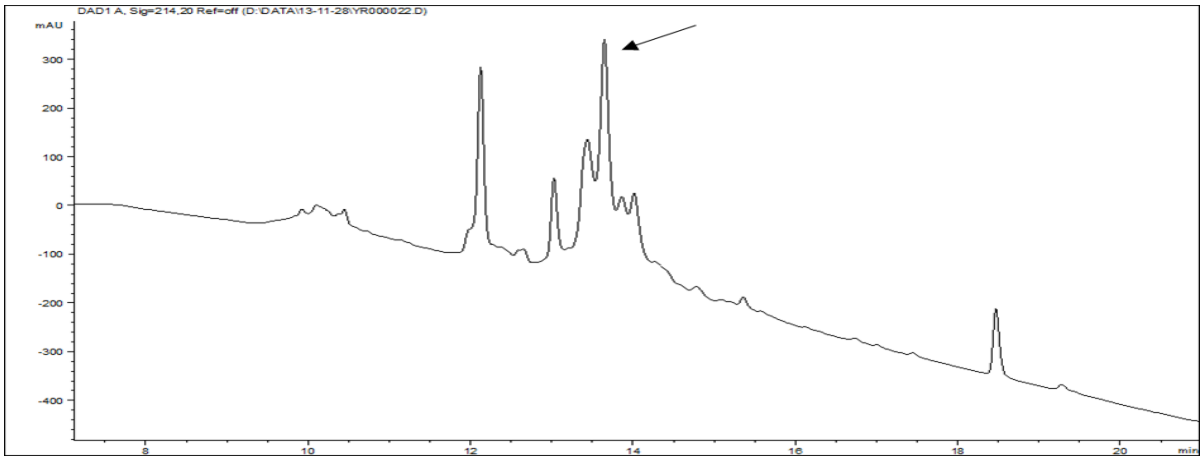


Figure 11.2.47: RP-HPLC Chromatogram of crude compound **11.2.33** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

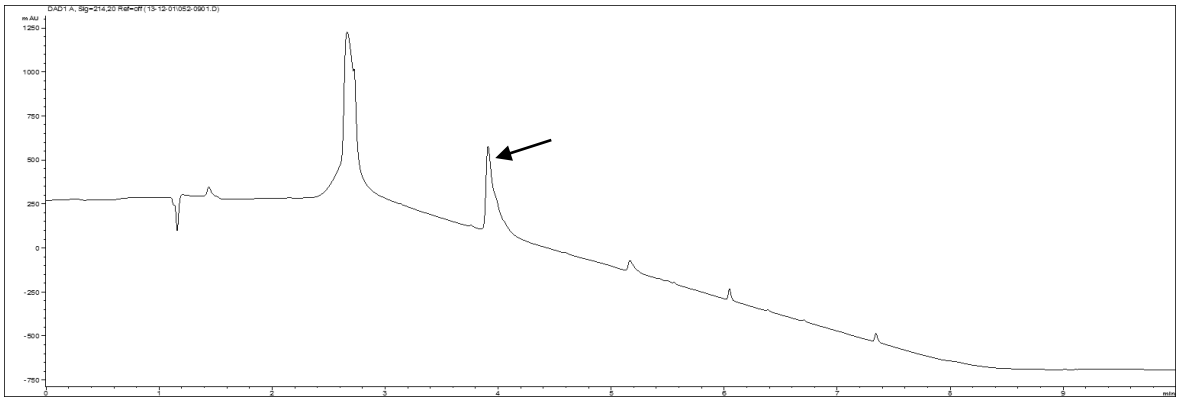


Figure 11.2.48: RP-HPLC trace of compound **11.2.33** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

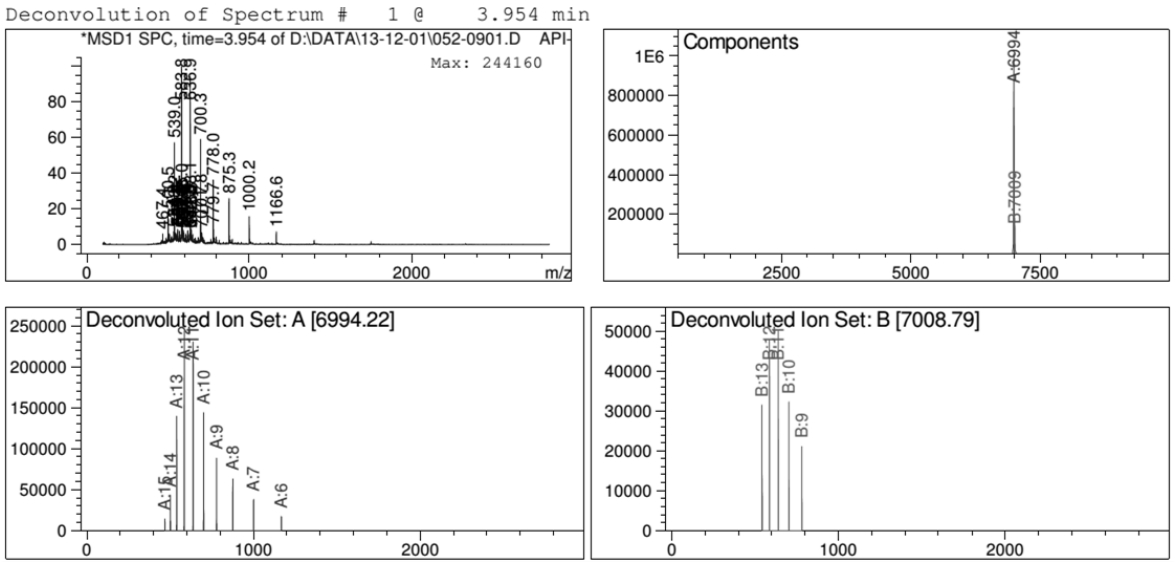


Figure 11.2.49: ESI-MS from LC-MS at t = 3.954 min for pure compound **11.2.33**. E.M calcd. for $C_{311}H_{484}N_{102}O_{84}$ = 6991.67 and E.M. + NH_4^+ = 7009.67 and deconvoluted mass found 6994.22 and 7008.79 respectively.

Synthesis of 11.2.34: Scaffold **11.2.3** (3.11 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.24** (3.85 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.51 mg, 12 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.

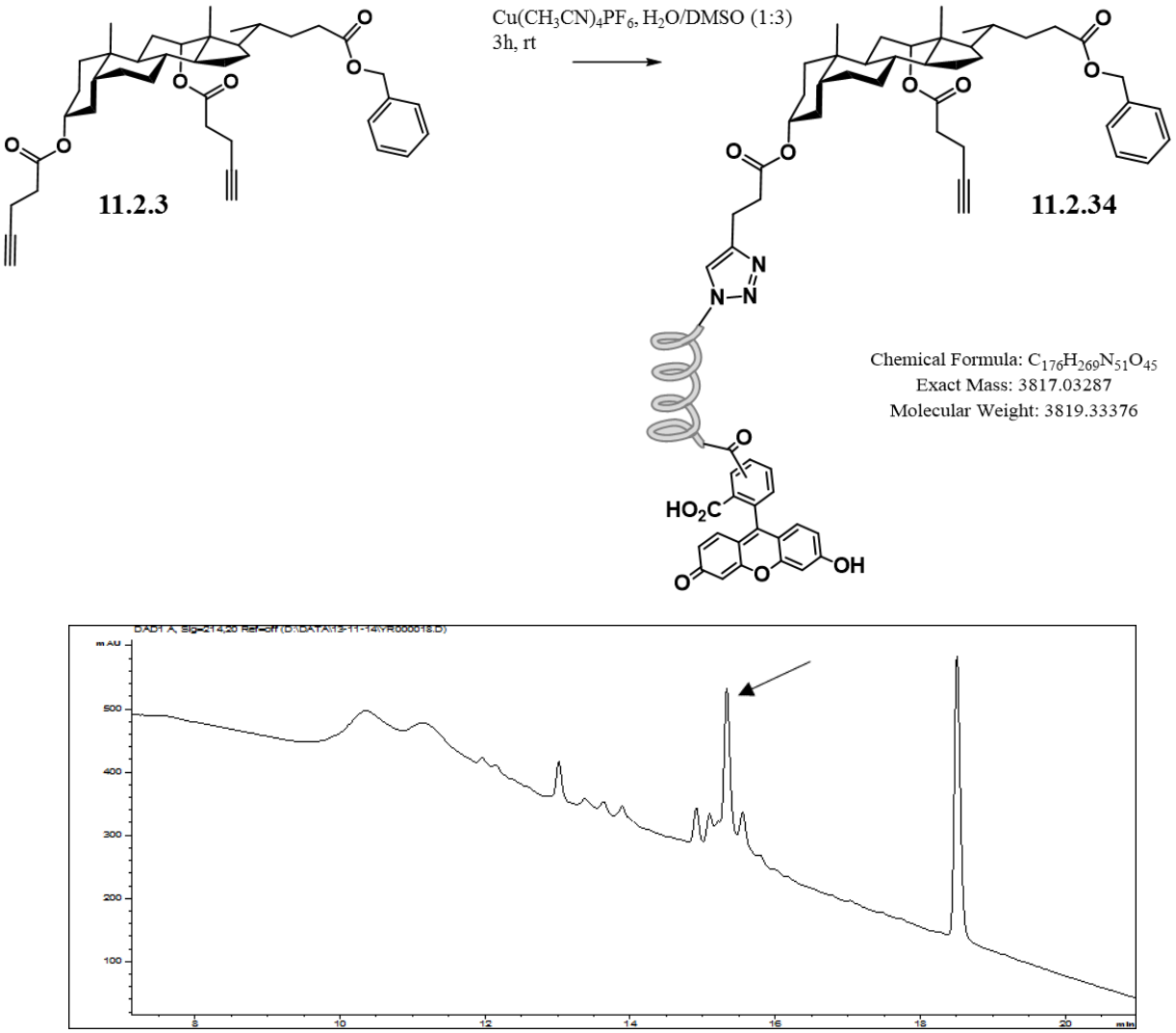


Figure 11.2.50: RP-HPLC Chromatogram of crude compound **11.2.34** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

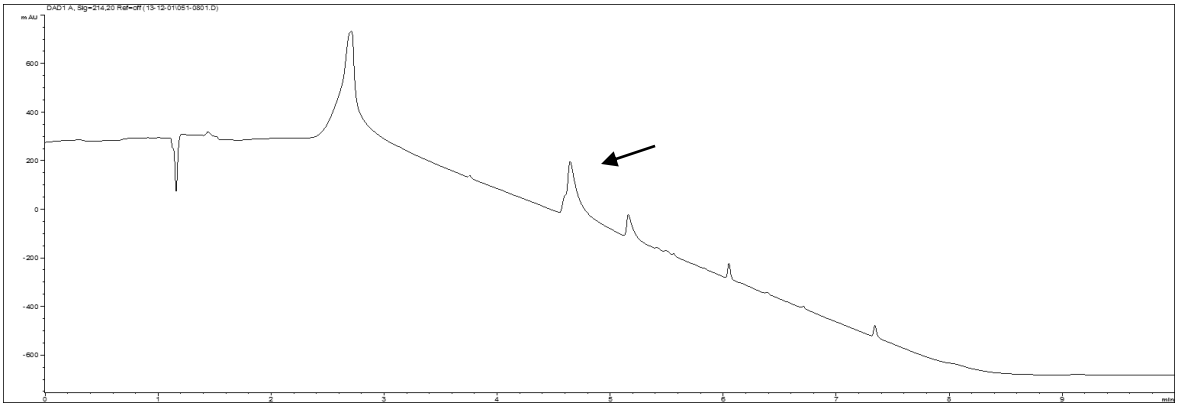


Figure 11.2.51: RP-HPLC trace of compound **11.2.34** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

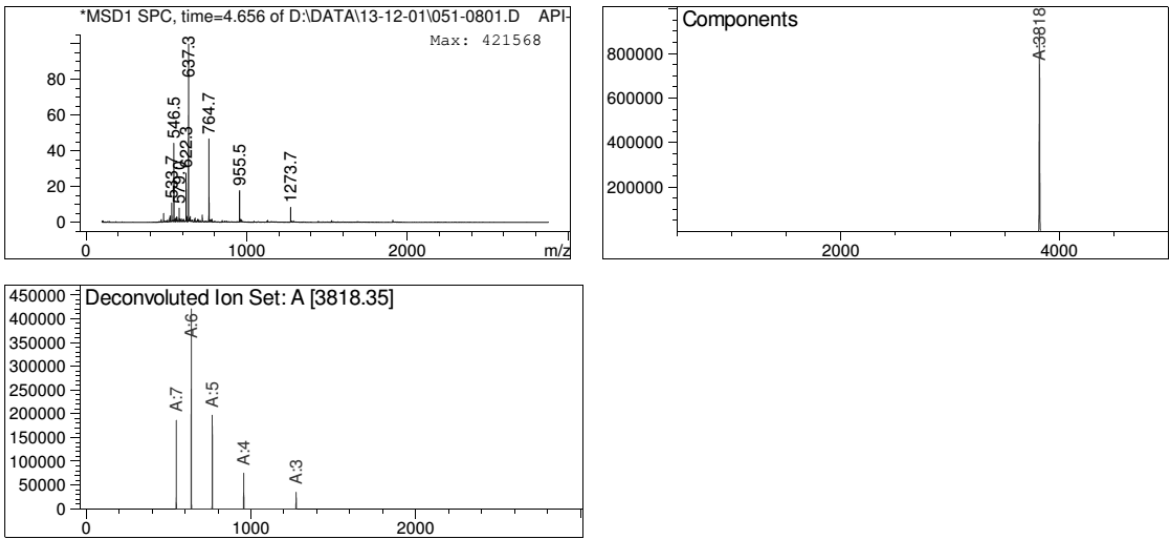


Figure 11.2.52: ESI-MS from LC-MS at t = 4.656 min for pure compound **11.2.34**. E.M calcd. for $C_{176}H_{269}N_{51}O_{45}$ = 3817.03 and deconvoluted mass found 3818.35.

11.2.6 EMSA studies I

All gels have been run according to the protocols described in sections 10.4 & 10.5 (depending on whether non-radiolabelled or radiolabelled DNA was used):

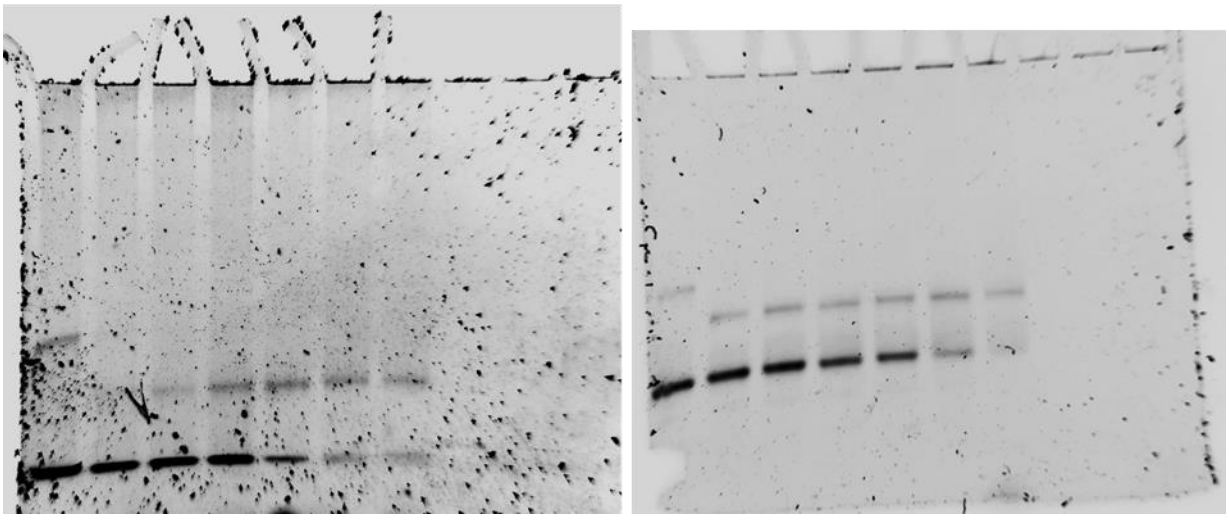


Figure 11.2.53. (left) EMSA titration of the dipodal peptidosteroid **11.2.29** to the Random dsDNA. Lane 1-10 contain respectively 0, 0.04, 0.12, 0.17, 0.20, 0.22, 0.26, 0.30, 0.35, 0.44 µM and compound **11.2.30** (right) lanes 1-10 contains respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of the mimic, stained with SybrGold.

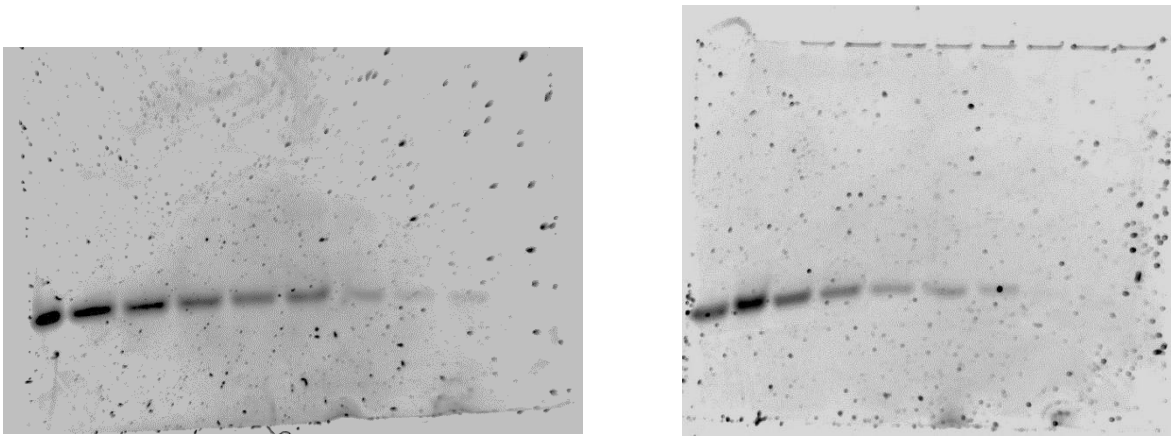


Figure 11.2.54. EMSA titration of the dipodal peptidosteroid **11.2.29** (left) and **11.2.30** (right) to the Random dsDNA. Lane 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 μM of the mimic, stained with SybrGold.

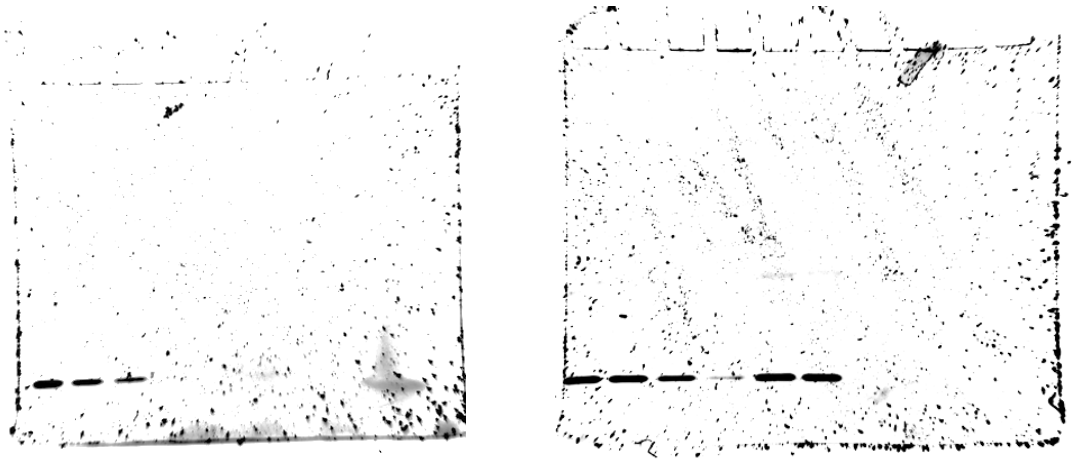


Figure 11.2.55. EMSA titration of the dipodal peptidosteroid **11.2.31** (left) and **11.2.32** (right) to the Random dsDNA. Lanes 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 μM of the mimic, stained with SybrGold.

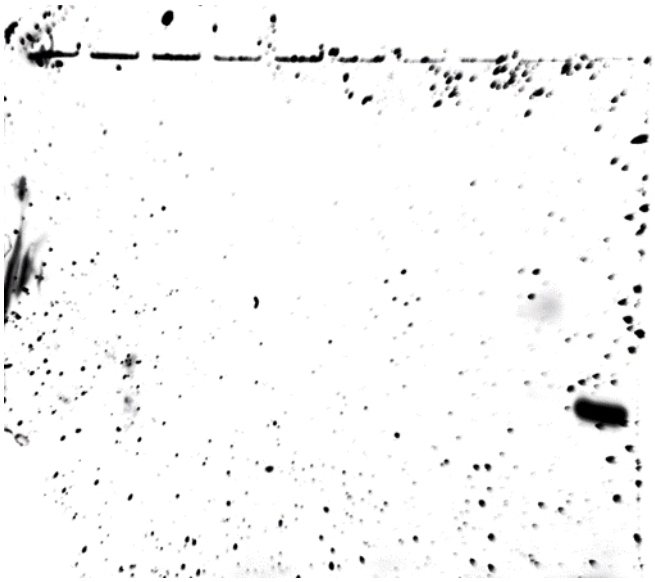


Figure 11.2.56. EMSA titration of the dimeric GCN4 **11.2.28** to the CRE binding site. Lane 1-10 contain respectively 0, 0.25, 0.5, 0.68, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5 µM of the mimic, stained with SybrGold.

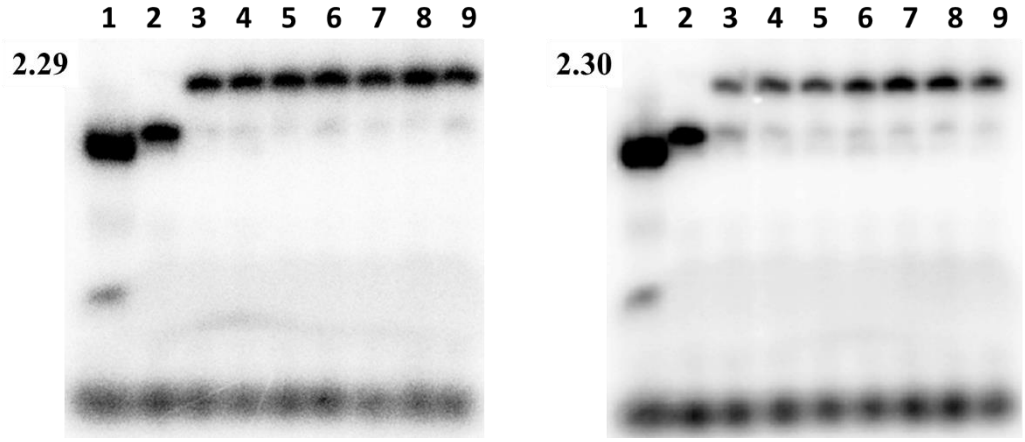


Figure 11.2.57. EMSA titration of the dipodal peptidosteroid conjugates **2.29** and **2.30** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 5nM: First lane in all the gels: pyrimidine strand. Lanes 2-9 contain peptide concentrations of 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 µM for **2.29** and **2.30**.

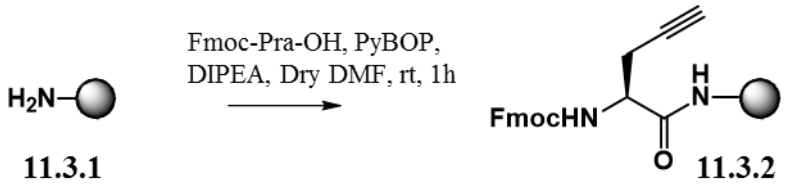
CHAPTER 11.3

11.3.1 Peptide synthesis

Attachment of the first residue was performed manually. Glycine and GABA were coupled as second and third residues respectively to increase the space between the scaffold and the peptide. Residues D226-Q248 were coupled by automated peptide synthesis using HBTU as coupling reagent and 20% piperidine/NMP for Fmoc deprotection. As a last residue and to increase UV-absorption and facilitate HPLC analysis, 4-acetamidobenzoic acid was coupled to the N-terminus. Cleavage and deprotection of the peptide was performed with a cocktail mixture containing TFA/TIS/H_2O (95/2.5/2.5) for 4 hours at room temperature. Peptide was obtained by precipitation with cold ether and lyophilization. In view of the satisfying crude integrity of the peptide, no further purification efforts were undertaken at this stage.

Manual coupling protocols:

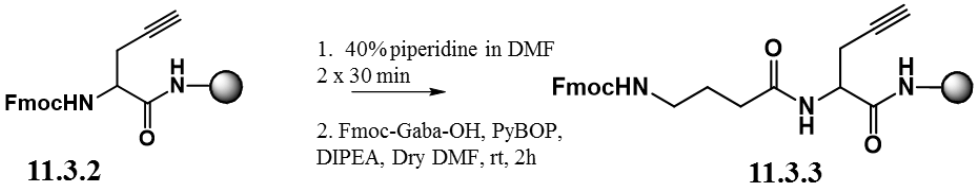
Immobilization of Fmoc-Pra-OH on Rink amide ChemMatrix resin



11.3.1 **11.3.2**

To a suspension of resin **11.3.1** (100 mg, 0.54 mmol/g) in DMF (dry, 10 mL/g resin), were added Fmoc-Pra-OH (54.5 mg, 0.162 mmol), PyBOP (84.3 mg, 0.162 mmol) and DIPEA (0.056 mL, 0.324 mmol). The mixture was shaken at room temperature for 1 h. After reaction, the resin was washed with DMF, ACN and MeOH. The resin was then capped with acetic anhydride (0.031 mL, 0.324 mmol) and DIPEA (0.056 mL, 0.324 mmol) in dry DMF (2 mL) 2 times for 30 min.

Fmoc deprotection and coupling of Fmoc-GABA-OH



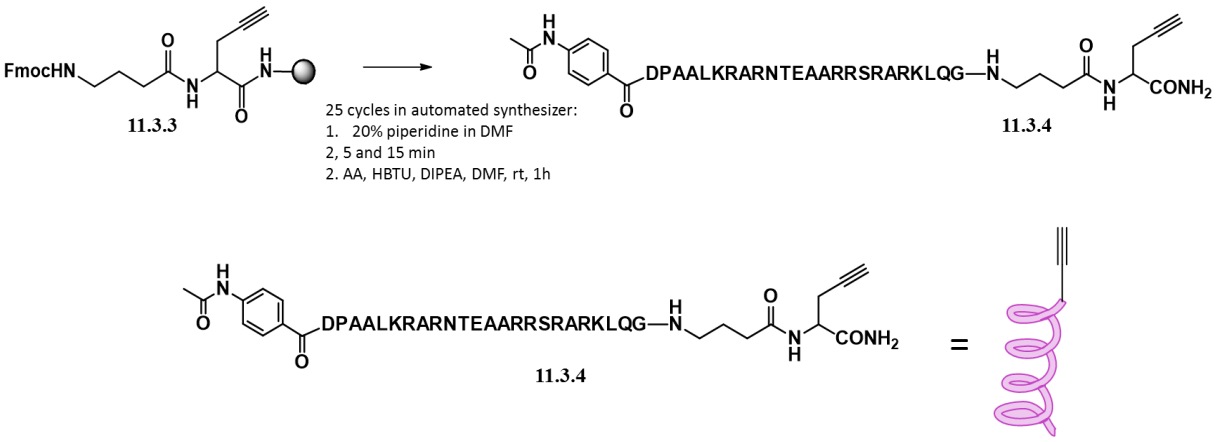
11.3.2 **11.3.3**

NHFmoc deprotection of **11.3.*2***. After an initial DMF washing step, resin **11.3.2** (0.054 mmol) was successively treated twice for 30 min with a piperidine solution in DMF (40 % v/v, 2 mL) at ambient temperature, applying intermediate filtration under reduced pressure and washing with DMF, while the final resin is additionally washed with ACN and DCM.

Coupling of Fmoc-GABA-OH. After Fmoc deprotection, Fmoc-GABA-OH (70.2 mg, 0.216 mmol, 0.5 M), PyBOP (112.4 mg, 0.216 mmol, 0.5 M) and DIPEA (0.038 mL, 0.216 mmol, 2 M) were added to a

suspension of resin in dry DMF (3 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

Automated synthesis of linear peptide Aba-DPAALKRARNTEAARRDRARKLQG-Gaba-Pra-$CONH_2$



Automated solid phase peptide synthesis was carried out on a Syro synthesizer from Biotage using standard Fmoc/tBu chemistry with HBTU as coupling reagent and 20% piperidine in NMP as deprotection reagent.

Fmoc protected resin (0.054 mmol) was subjected to automated synthesis where solutions of Fmoc-Nα-protected amino acids (0.5 M in NMP) were prepared. Each coupling reaction lasted 1h (amino acids 0.250 mmol; HBTU 0.250 mmol in DMF, 0.5 M; DIPEA 0.250 mmol 2 M) and was followed by Fmoc deprotection with 40% piperidine in NMP. After each reaction the resin was washed with NMP (9 x). The peptide was then cleaved for the resin and deprotected with a cocktail of TFA/TIS/water = 95:2.5:2.5 for 4h. After precipitation in cold ether, the peptide was analyzed by RP-HPLC and MALDI.

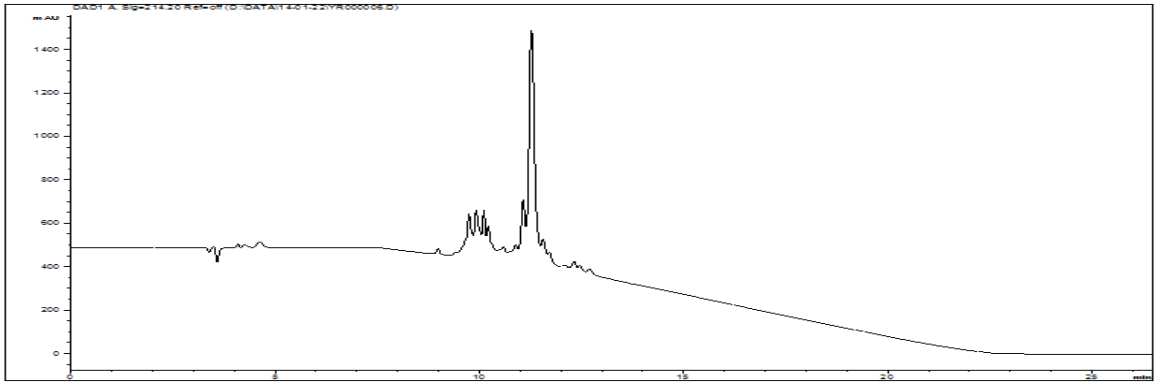


Figure 11.3.1: RP-HPLC Chromatogram of crude compound **11.3.4** (Jupiter C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

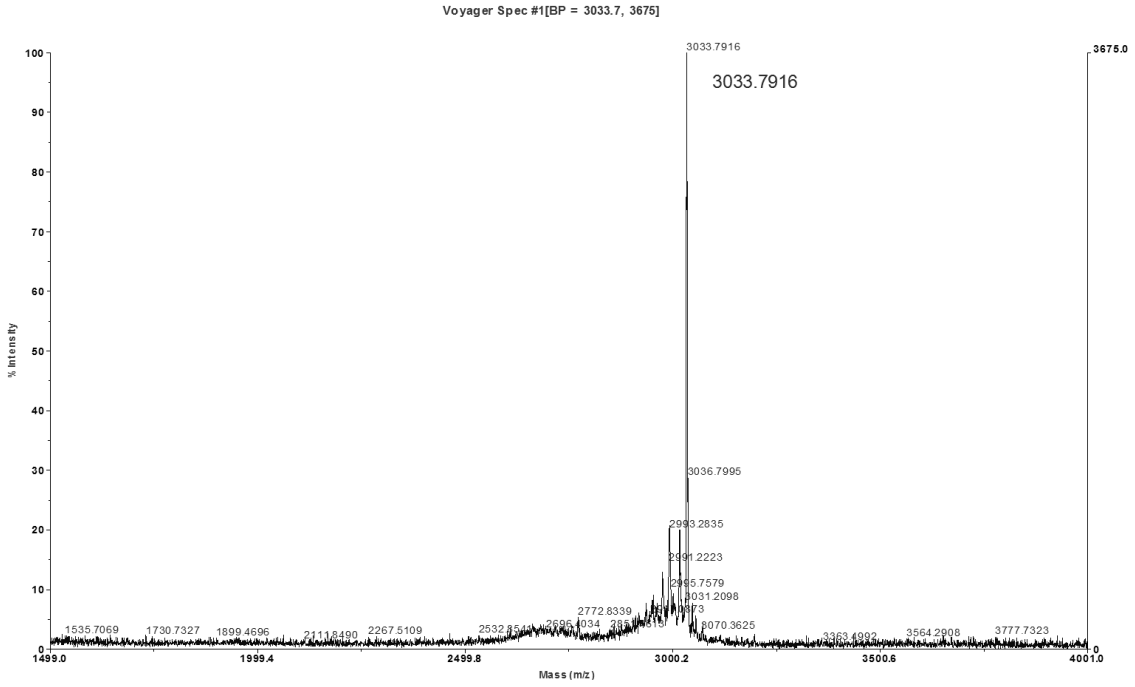


Figure 11.3.2: MALDI-TOF Spectrum of crude compound **11.3.4**. Calculated m/z (100%) = 3032.68; m/z (91.8%) = 3033.68. Found: 3033.7916 = M + H^+.

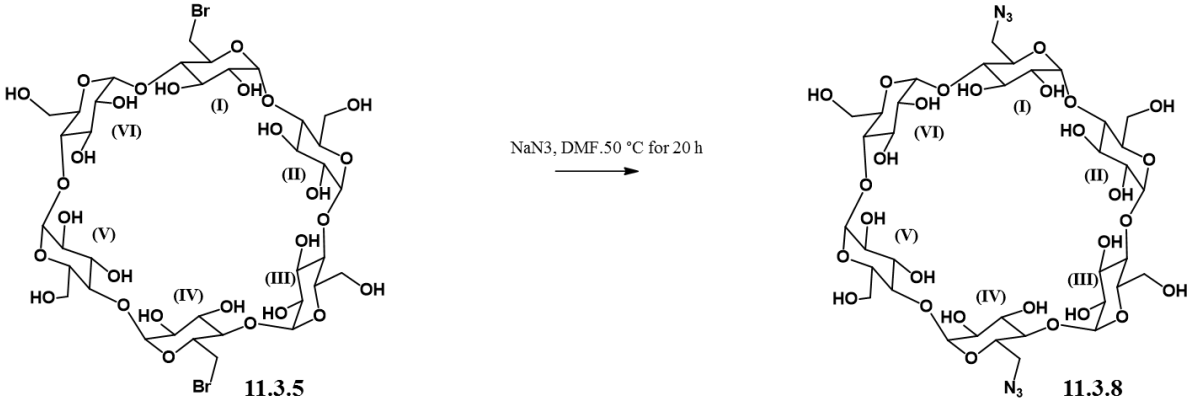
11.3.2 Synthesis of Cyclodextrin Scaffolds

Complete synthesis and characterization has been performed by the group of Prof. Tomas Kraus (to be published). A part of the synthesis is described here.

2. I Cyclodextrin template molecules:

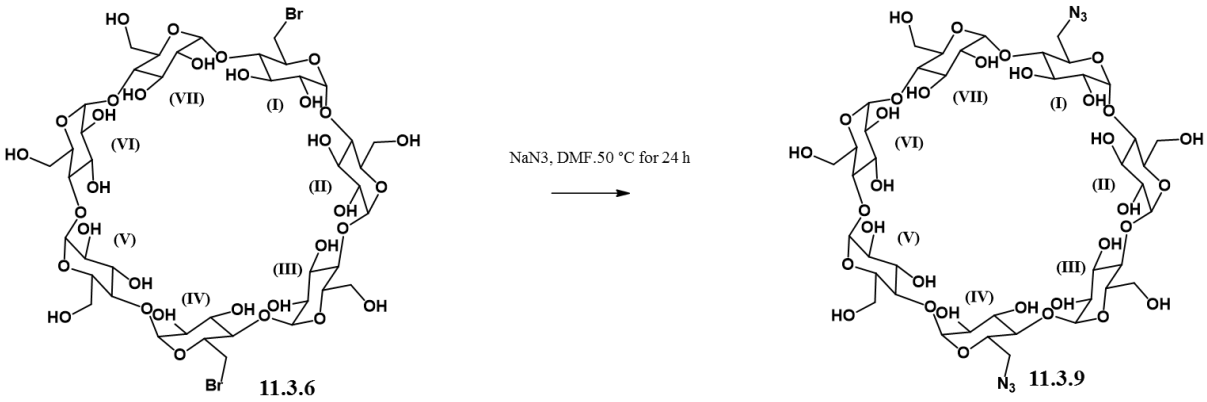
General experimental procedures: Preparative reversed–phase (RP) chromatography was carried out using medium pressure columns containing C-18 modified silica (Phenomenex Luna, 15 µm). Thin-layer (TLC) and reversed-phase thin-layer chromatography (RPTLC) were performed with precoated Silica Gel 60F and RP-18 F plates (E. Merck) respectively, which were visualized by spraying with an aqueous solution of phosphomolybdenic acid containing 5% of H_2SO_4 and heating. All chemicals used were commercially available. Compounds **11.3.8, 11.3.9** and **11.3.10** were prepared according to the known procedure[109,263,264]. Satisfactory elemental analysis could not be obtained for hydrophilic compounds 1 – 3 unless variable numbers of water molecules were taken into account. Thus, calculations based on weights of these compounds (molarity, yield) are related to the hydrated molecules.

$6^I,6^{IV}$-Dideoxy-$6^I,6^{IV}$-diazido-α-cyclodextrin **11.3.8**



To a solution of $6^I,6^{IV}$-dideoxy-$6^I,6^{IV}$-dibromo-α-cyclodextrin trihydrate[1] **11.3.5** (0.367 g, 0.318 mmol) in DMF (5 ml) sodium azide (0.130 g, 2 mmol) was added. The mixture was stirred at 50 °C for 20 h under argon atmosphere and then the solvent was evaporated under reduced pressure. The solid material was dissolved in a warm mixture of methanol-water (8:2, 5 ml) and charged onto RP column. Gradient elution with a mixture of methanol-water from 1:9 to 4:6 followed with evaporation of selected fractions and drying of the isolated material over phosphorus pentoxide *in vacuo* gave compound **11.3.8** as white amorphous material (0.305 g, 90%, calcd. for dihydrate). R_f = 0.3 (RP-TLC, methanol-water 1:1). Elemental analysis: Calcd. for $C_{36}H_{58}N_6O_{28}$. $2H_2O$: C, 40.83, H 5.90; N 7.94; Found: C, 40.70, H, 5.77, N 7.63. ESI-MS calcd for [M + Na^+]: 1045.3; found: 1045.3. ^{1}H NMR (see Table 11.3.1); ^{13}C NMR (see Table 11.3.2).

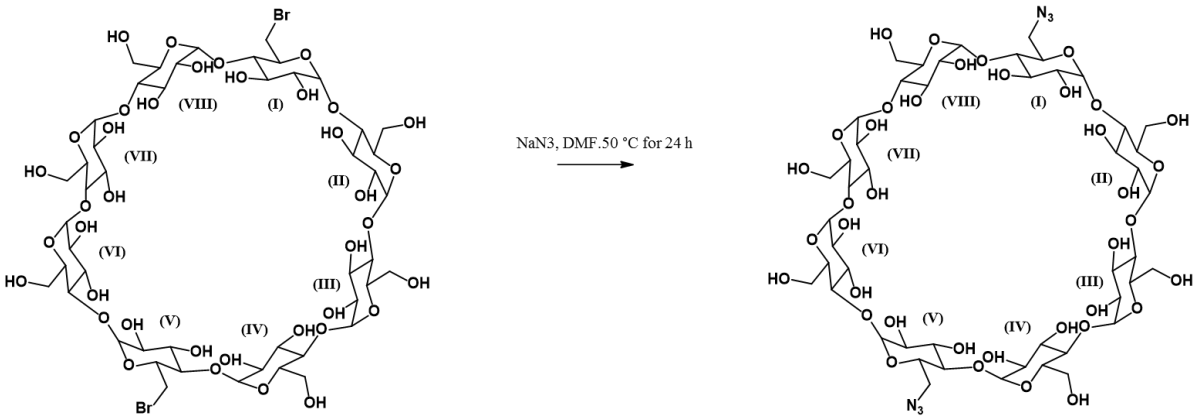
$6^I,6^{IV}$-Dideoxy-$6^I,6^{IV}$-diazido-β-cyclodextrin **11.3.9**.



To a solution of $6^I,6^{IV}$-dideoxy-$6^I,6^{IV}$-dibromo-β-cyclodextrin tetrahydrate[2] **11.3.6** (0.392 g, 0.294 mmol) in DMF (8.5 mL) sodium azide (0.115 g, 1.76 mmol) was added and the mixture was stirred at 50 °C for 24 h under argon atmosphere. Then the solvent was evaporated under reduced pressure to about a half of the volume and acetone (50 mL) was added under stirring. The precipitate, which formed upon addition of acetone, was collected on a glass frit, washed with acetone and dried *in vacuo*. Subsequently, it was dissolved in a warm mixture of methanol-water (9:1, 5 ml) and charged onto RP column. Gradient elution

with a mixture of methanol-water from 1:9 to 4:6 followed with evaporation of selected fractions gave compound **11.3.9** (320 mg, 83 %, calcd as heptahydrate) as white amorphous material after drying the isolated material over phosphorus pentoxide *in vacuo*. R_f = 0.3 (RP-TLC, methanol-water 1:1). Elemental analysis: Calcd for $C_{42}H_{68}N_6O_{33}$. $7H_2O$: C, 38.48; H, 6.30; N, 6.41; Found: C, 38.35, H, 6.48, N 5.60. ESI-MS calcd for [M+Na^+]: 1207.4; found 1207.4. ^{1}H NMR (see Table 11.3.1); ^{13}C NMR (see Table 11.3.2).

6^I,6^V-Dideoxy-6^I,6^V-diazido-γ-cyclodextrin **11.3.10.**



11.3.7 **11.3.10**

To a solution of 6^I,6^V-dideoxy-6^I,6^V-dibromo-γ-cyclodextrin pentahydrate[3] **11.3.7** (43 mg, 28.4 μmol) in DMF (1 ml) sodium azide (11 mg, 0.171 mmol) was added. The mixture was stirred at 50 °C for 24 h under argon atmosphere and then solvents were evaporated under reduced pressure. The residue was dissolved in a mixture of methanol-water (8:2, 3 ml) and charged onto RP column (10 g, Phenomenex Luna 15 μm). Gradient elution with a mixture of methanol-water from 1:9 to 4:6 followed with evaporation of selected fractions and drying of the isolated material over phosphorus pentoxide *in vacuo* gave compound **11.3.10** as white amorphous material (32 mg, 77%, calcd for hexahydrate). R_f = 0.4 (RP-TLC, methanol-water 4:6). Elemental analysis: Calcd for $C_{48}H_{78}N_6O_{38}$. $6H_2O$: C, 39.62, H 6.23, N 5.78; Found: C, 39.73, H, 5.91, N 5.39. HR-MALDI-MS calcd for [M+Na^+]: 1369.4248; found 1369.4266. ^{1}H NMR (see Table 11.3.1); ^{13}C NMR (see Table 11.3.2).

2. II NMR spectra of cyclodextrin template molecules

The NMR spectra of compounds **11.3.8, 11.3.9** and **11.3.10** were measured on Bruker AVANCE-600 instrument (^{1}H at 600.13 MHz and ^{13}C at 150.9 MHz) with a cryoprobe in CD_3OD at 25 °C. Spectra were referenced to solvent peak (using δ_H (CHD_2OD) = 3.31 ppm; δ_C (CD_3OD) = 49.0 ppm). Structural assignment of proton and carbon signals was achieved combining 1D-^{1}H and ^{13}C-spectra with homonuclear 2D-H,H-COSY, 2D-H,H-TOCSY, 2D-H,H-ROESY and heteronuclear 2D-H,C-HSQC and 2D-H,C-HMBC spectra. The ^{1}H and ^{13}C NMR data are summarized in Tables S1 and S2. Complete sequential structural assignment of proton and carbon signals has been achieved only for "symmetrical" α-cyclodextrin derivative **11.3.8**. For β- and γ-cyclodextrin derivatives **11.3.9** and **11.3.10**, respectively, only partial structural assignment was possible due to the poor separation of signals. The sequential positions of glucose

units in Table S1 and S2 are labeled with numbers I to VIII as it is shown in Figure S1. All observed coupling constants *J*(H,H) of ring protons of glucose units (not given in Table 1) showed the values characteristic for $^4C_1$ conformation (*J*(1,2) = 3.5 – 4.0 Hz, *J*(2,3) ~ *J*(3,4) ~ *J*(4,5) = 8.5 – 10.0 Hz).

Table 11.3.1. Proton chemical shifts of compounds **11.3.8-10** in CD_3OD at 600.13 MHz

Comp.	Glucose residue	H-1 (d)	H-2 (dd)	H-3 (dd)	H-4 (dd)	H-5 (ddd)	H-6a (dd)	H-6b (dd)
11.3.8	I, IV	4.971	3.509	3.916	3.447	3.961	3.728	3.612
	II, V	4.940	3.494	3.935	3.517	3.821	~3.882 (2H)	
	III, VI	4.907	3.494	3.911	3.469	3.793	~3.833 (2H)	
11.3.9	I	4.994	3.53-3.50 (7H)	3.88-3.83 (7H)	3.456	3.891 (2H)	3.713	3.590
	IV	4.994			3.450		3.713	3.580
	II	4.975			3.428	3.76-3.71 (5H)	3.88 – 3.83 (10H)	
	III	4.968			3.424			
	V	4.962			3.51 (3H)			
	VI	4.943						
	VII	4.937						
11.3.10	I, V	5.032	3.528	3.826	3.424	3.878	3.695	3.568
	II, VI	5.013	3.516	3.839	3.510	3.742	3.80-3.89 (12H)	
	III, VII	5.004	3.516	3.833	3.493	3.738		
	IV, VIII	4.978	3.522	3.817	3.462	3.705		

[a] Chemical shift values in columns may be interchanged (except for H-6a,H-6b).

Table 11.3.2. Carbon-13 chemical shifts of compounds **11.3.8-10** in CD_3OD at 150.9 MHz

Comp.	Glucose residue	C-1	C-2	C-3	C-4	C-5	C-6
11.3.8	I, IV	103.43	73.84	74.92	84.36	72.50	52.82
	II, V	103.67	73.82	75.15	83.38	73.75	61.93
	III, VI	103.71	73.76	75.23	83.40	73.89	61.99

11.3.9	I	103.99	74.21	74.82 (2C)	84.26	73.85	52.76
	IV	103.97	74.17 (3C)		84.21	73.81	52.73
	II	103.92		74.77	83.23	73.70	61.95 (2C)
	III	103.91		74.68	83.20	73.67	
	V	103.84	74.14	74.65	83.04 (2C)	73.63	61.87 (2C)
	VI	103.53 (2C)	74.09	74.55		72.42 (2C)	
	VII		74.07	74.52	83.02		61.84
11.3.10	I, V	103.44	74.42 (2C)	74.50	83.95	72.49	52.71
	II, VI	103.84		74.56	82.64	73.79	61.99
	III, VII	103.90	74.47	74.62	82.82	73.74	61.92 (2C)
	IV, VIII	104.02	74.50	74.65	82.87	73.95	

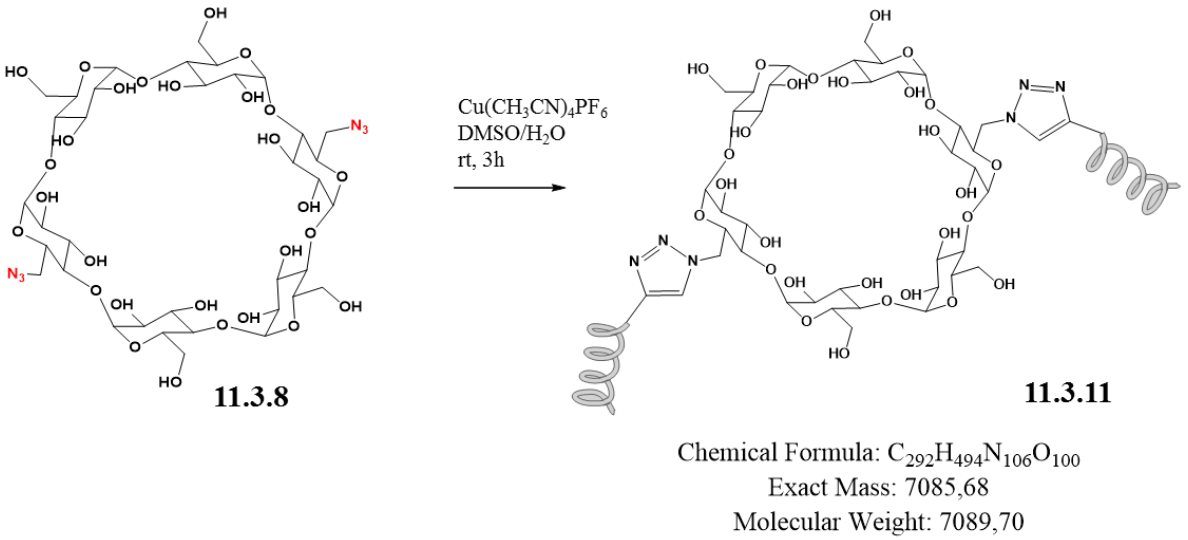
[a] Chemical shift values in columns may be interchanged (except for C-6).

11.3.3 CuAAC Conjugation II: peptides & cyclodextrins

Peptide-Cyclodextrin conjugation protocol via CuAAC

Using alfa cyclodextrin

Synthesis of 11.3.11: $6^{I},6^{IV}$-Dideoxy-$6^{I},6^{IV}$-diazido-α-cyclodextrin **11.3.6** (6.75 mg, 6.6 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.3.4** (5 mg, 1.6 µmol) was dissolved in 0.1 mL milliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (6.15 mg, 16 µmol) was dissolved in 75 µL dry DMSO and added to the reaction mixture. The reaction was stirred for 3h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC to obtain compound **11.3.11**. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



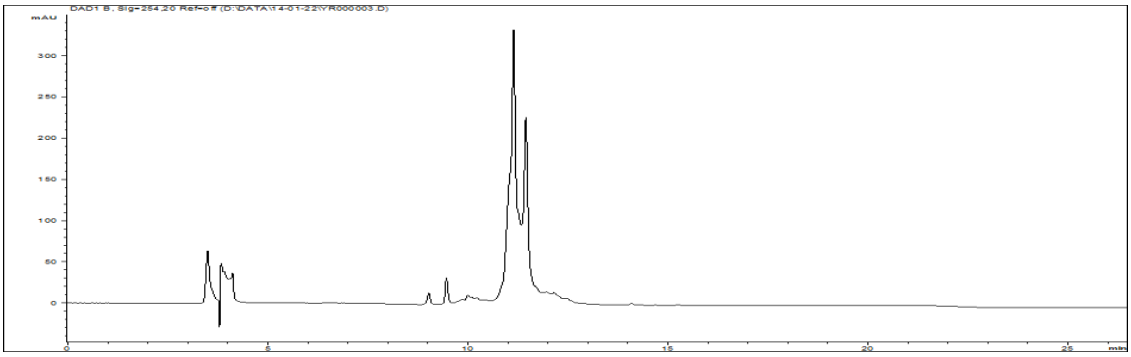
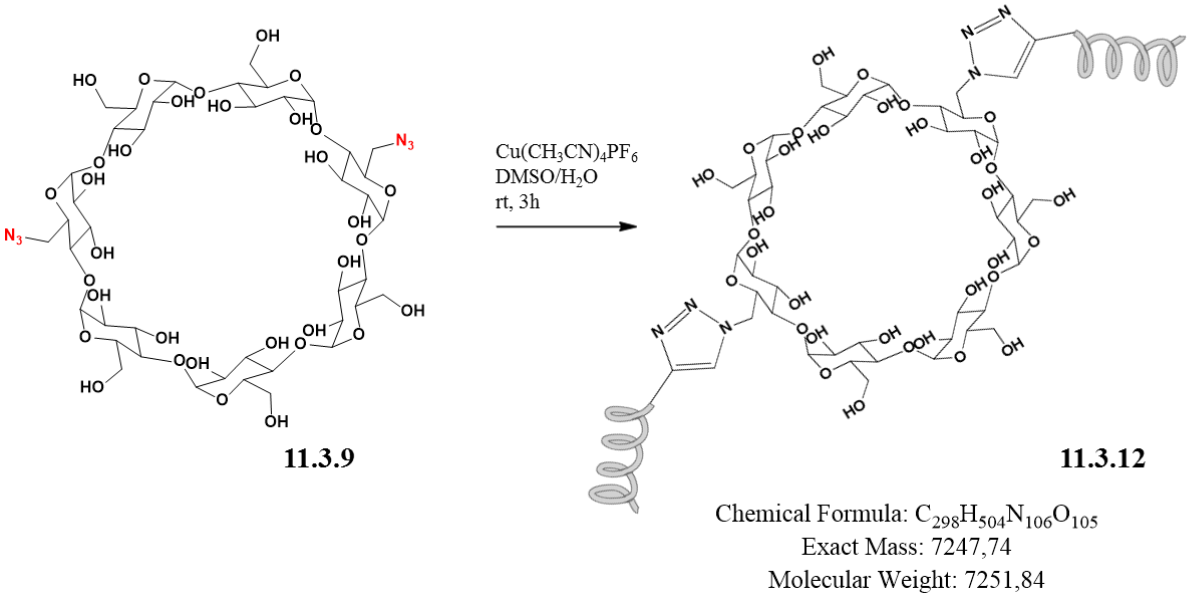


Figure 11.3.3: RP-HPLC trace (0-100% in 15 min on Jupiter C4, 300 Å) of the reaction mixture after 3 h.

The chromatogram shows two main peaks corresponding to the unreacted peptide (11.1 min) and desired conjugate (11.5 min). The sharpness of the peak of the conjugate allowed for easy purification by fraction collection.

Using beta cyclodextrin

Synthesis of 11.3.12: 6^{I},6^{IV}-Dideoxy-6^{I},6^{IV}-diazido-β-cyclodextrin (6.25 mg, 5 µmmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.3.4** (4 mg, 1.3 µmol) was dissolved in 0.1 mL milliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.91 mg, 13 µmol) was dissolved in 75 µL dry DMSO and added to the reaction mixture. The reaction was stirred for 3h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC to obtain compound **11.3.12**. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



11.3.9 **11.3.12**

Chemical Formula: $C_{298}H_{504}N_{106}O_{105}$
Exact Mass: 7247,74
Molecular Weight: 7251,84

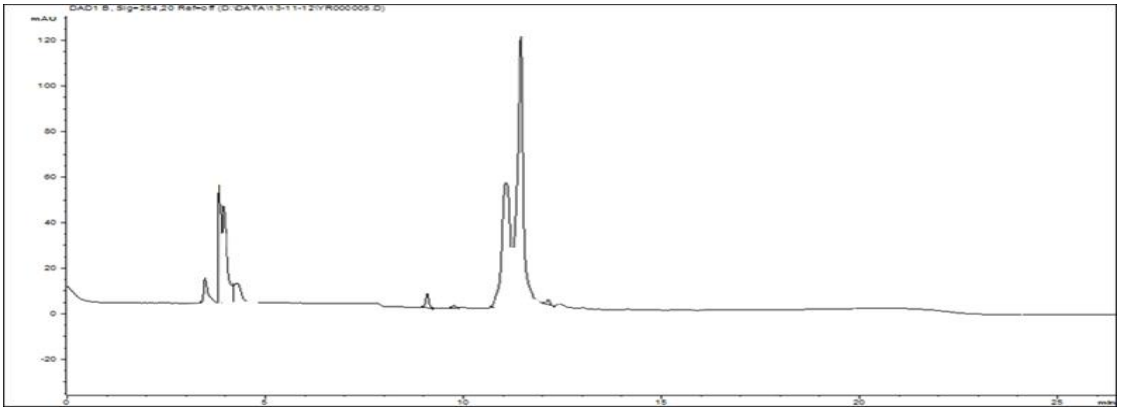
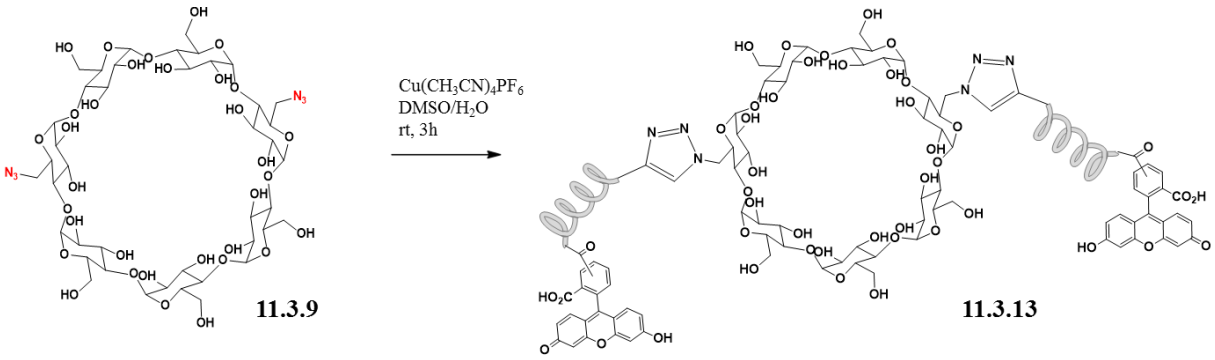


Figure 11.3.4: RP-HPLC trace (0-100% in 15 min on Jupiter C4, 300 Å) of the reaction mixture after 3 h. The chromatogram shows two main peaks corresponding to the unreacted peptide (11.1 min) and desired conjugate (11.4 min). The conjugation of the peptide to the β-cyclodextrin derivative proceeded better than to α and γ ones, as the major peak is the desired compound. The sharpness of the peak of the conjugate allowed for easy purification by fraction collection.

Synthesis of 11.3.13: $6^{I},6^{IV}$-Dideoxy-$6^{I},6^{IV}$-diazido-β-cyclodextrin (1 mg, 0.8 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.2.25** (8 mg, 2.5 µmol) was dissolved in 0.1 mL milliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (1.56 mg, 4 µmol) was dissolved in 75 µL dry DMSO and added to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC to obtain compound 33. Fractions were lyophilized and analyzed by LC-MS.



Chemical Formula: $C_{322}H_{510}N_{104}O_{113}$
Exact Mass: 7641.73582
Molecular Weight: 7646.12380

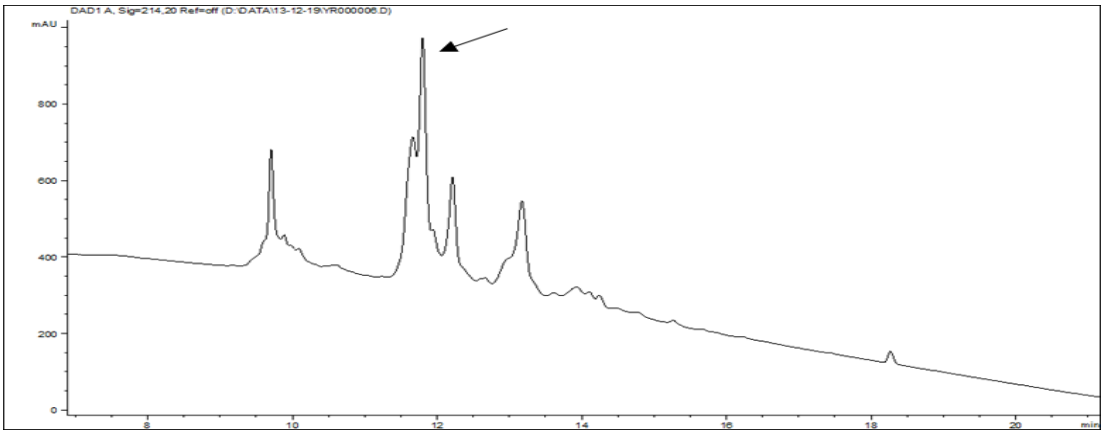
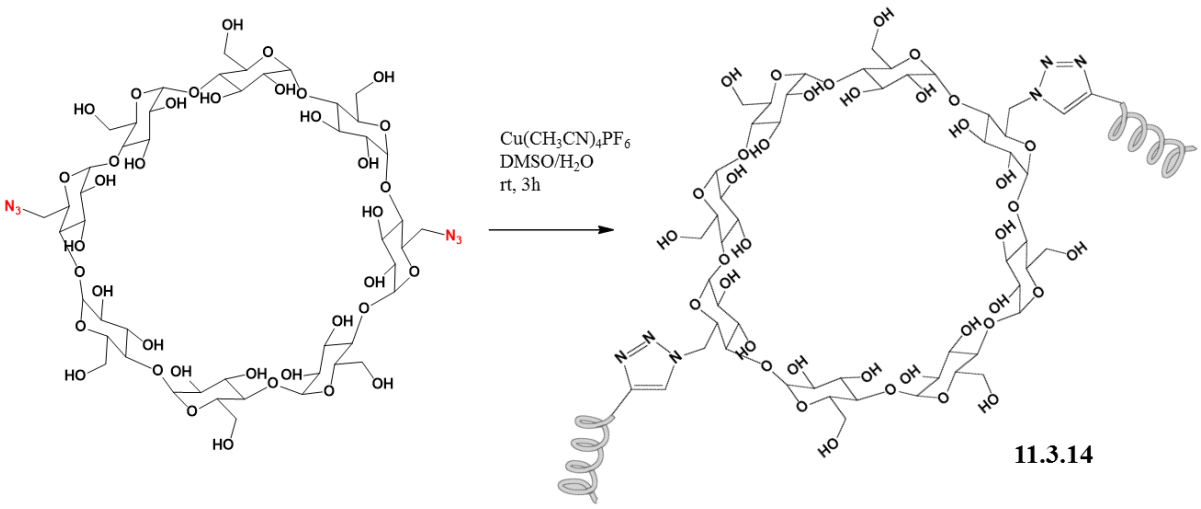


Figure 11.3.5: RP-HPLC trace (0-100% in 15 min on Jupiter C4, 300 Å) of the reaction mixture **11.3.13** after 3 h.

Using gamma cyclodextrin

Synthesis of 11.3.13.14: $6^{I},6^{V}$-Dideoxy-$6^{I},6^{V}$-diazido-γ-cyclodextrin (8.9 mg, 6.6 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11.3.4** (5 mg, 1.6 µmol) was dissolved in 0.1 mL milliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (6.15 mg, 16 µmol) was dissolved in 75 µL dry DMSO and added to the reaction mixture. The reaction was stirred for 3h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC to obtain compound **11.3.14**. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



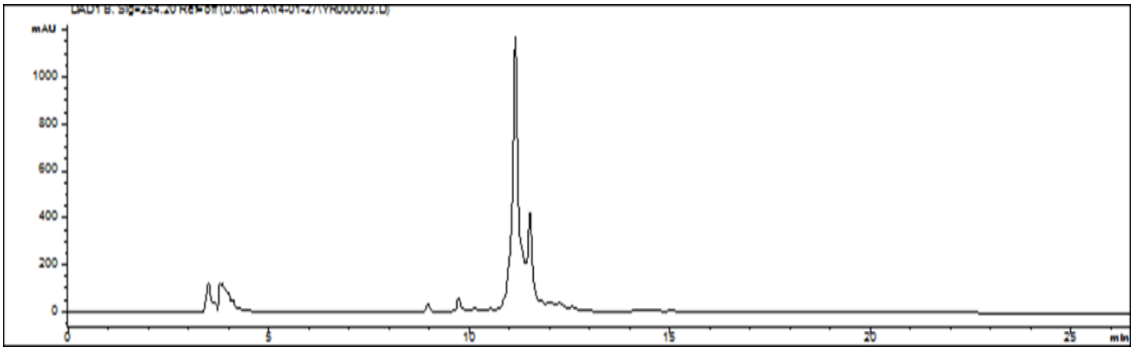


Figure 11.3.6: RP-HPLC (0-100% in 15 min on Jupiter C4, 300 Å) of the reaction mixture after 3 h.

The chromatogram shows two mean peaks corresponding to the unreacted peptide (11.1 min) and desired conjugate (11.5 min). The sharpness of the peak of the conjugate allowed for easy purification by fraction collection.

11.3.4 Analytical data of purified peptide-cyclodextrin conjugates

Compound **11.3.11**

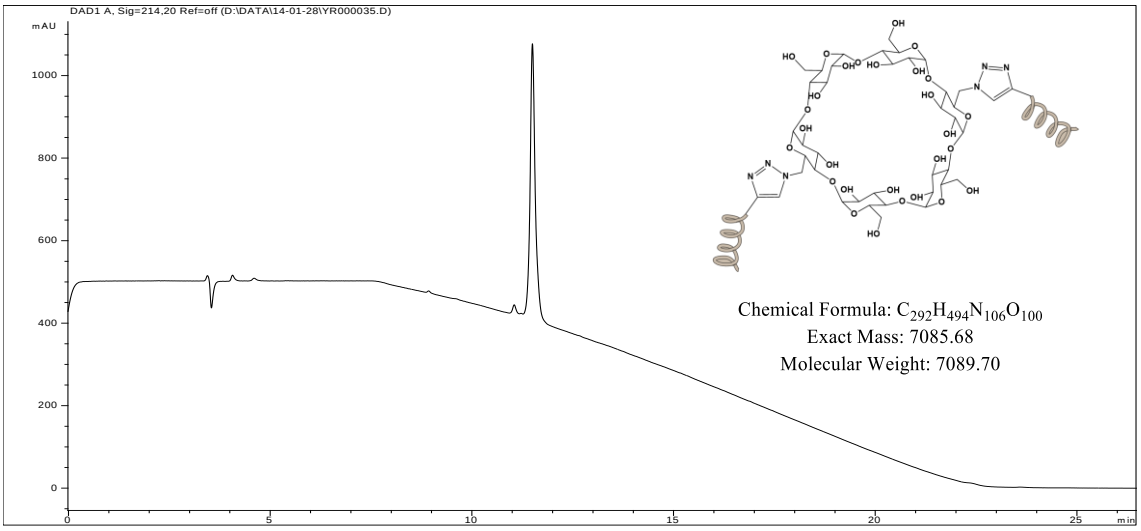


Figure 11.3.7: RP-HPLC Chromatogram of pure compound **11.3.11** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

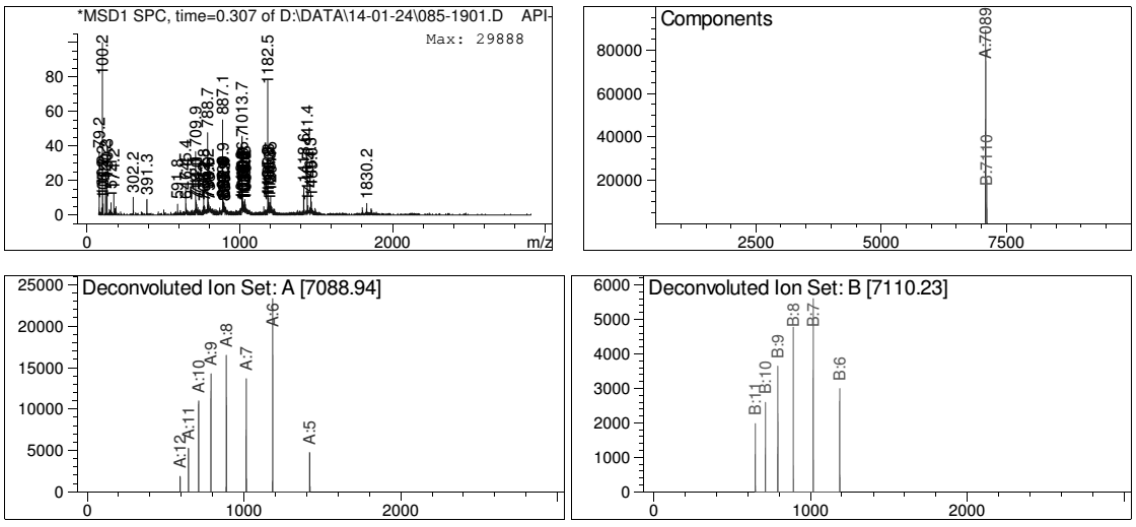


Figure 11.3.8: ESI-MS of compound **11.3.11**. Calcd. E.M. = 7085.68, M + Na^+ = 7108.68, Obsd. M = 7088.94, M + Na^+ = 7110.23.

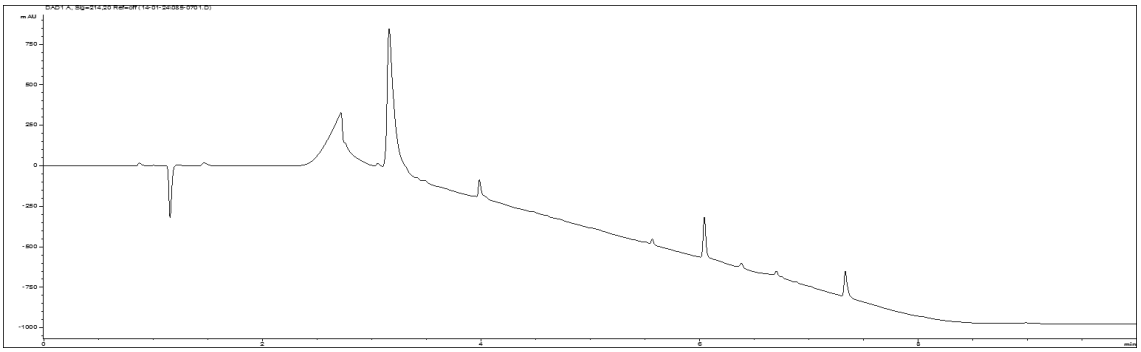


Figure 11.3.9: HPLC trace from LC-MS for compound **11.3.11** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

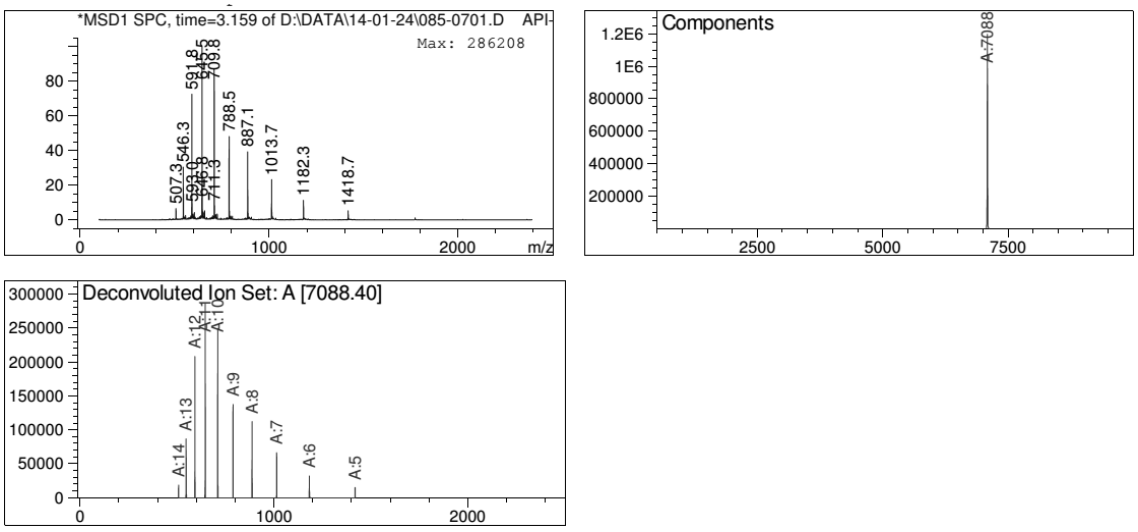


Figure 11.3.10: ESI-MS from LC-MS at r.t. = 3.159 min for purified compound **11.3.11**. E.M = 7085.68 and deconvoluted mass found 7088.40.

Compound **11.3.12** (Beta conjugate)

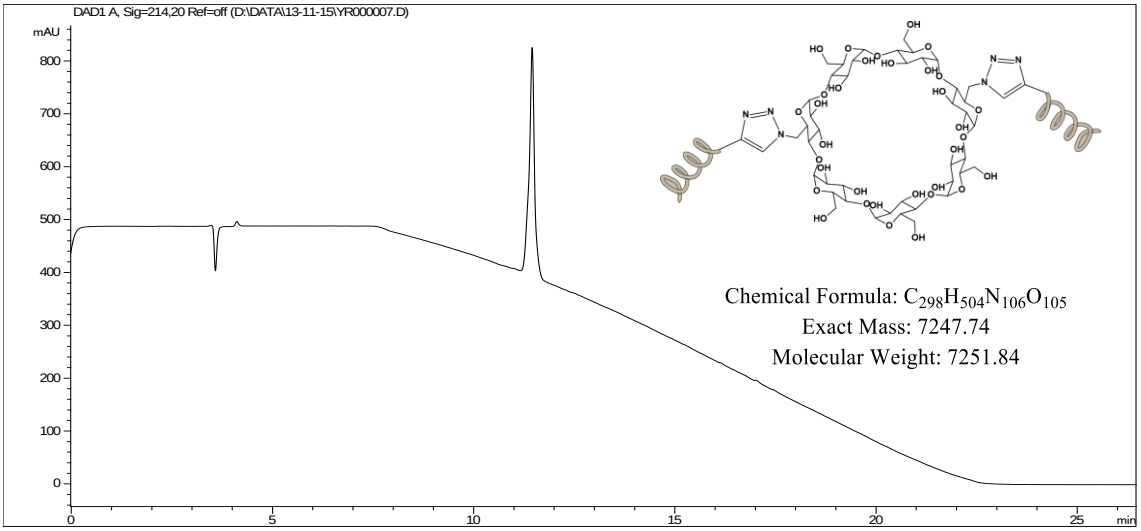


Figure 11.3.11: RP-HPLC Chromatogram of pure compound **11.3.12** (Jupiter C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

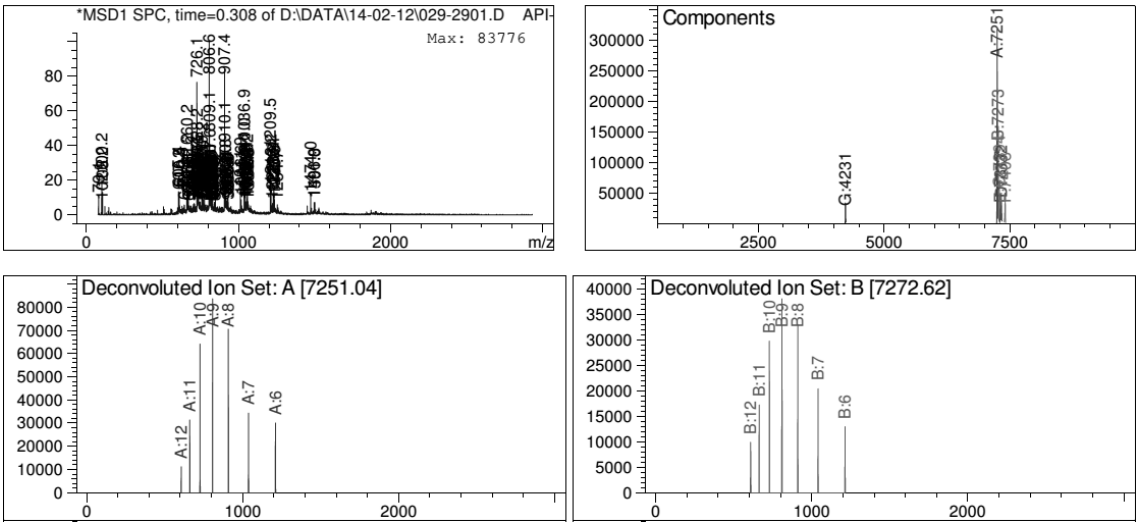


Figure 11.3.12: ESI-MS of compound **11.3.12**. Calcd. E.M. = 7247.74, M + Na^+ = 7270.74 and deconvoluted masses found 7251.04 and 7272.62 + other minor impurities.

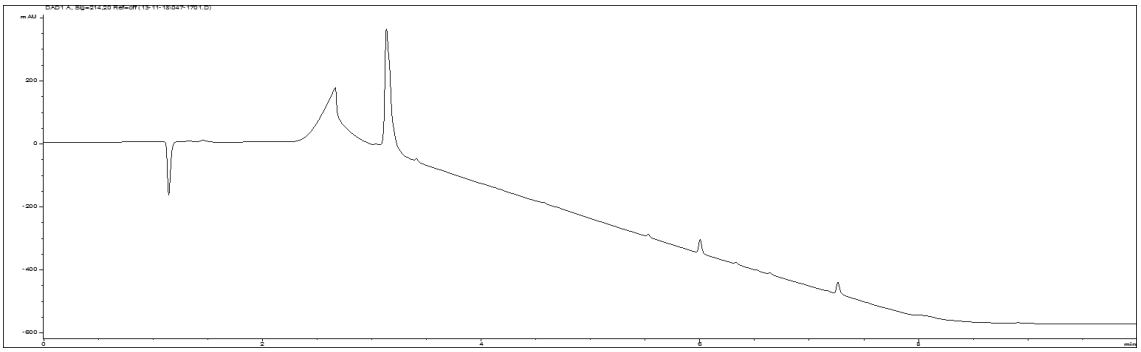


Figure 11.3.13: HPLC trace from LC-MS for compound **11.3.12** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

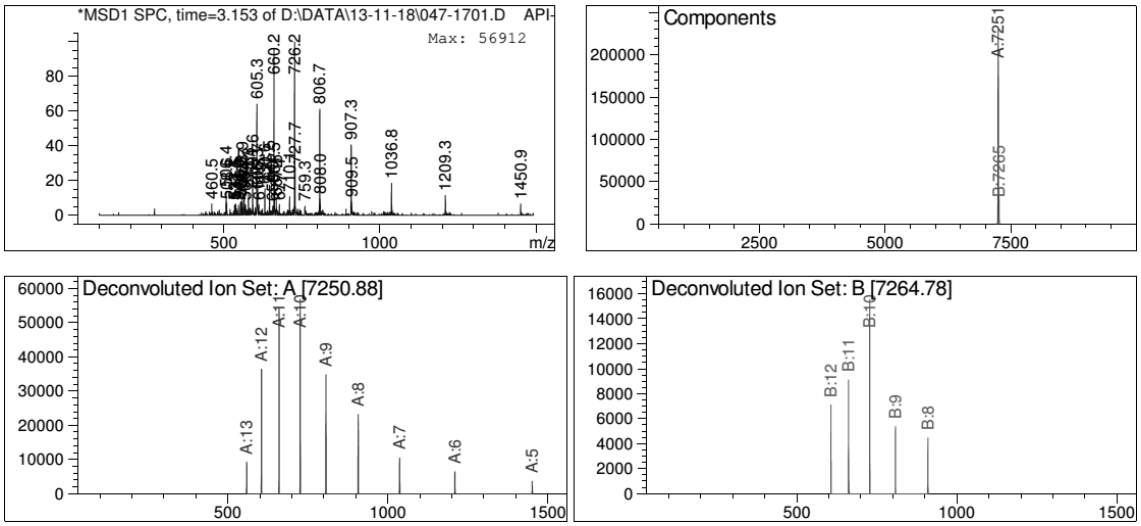


Figure 11.3.14: ESI-MS from LC-MS at r.t. = 3.153 min for purified compound **11.3.12**. E.M. = 7247.74, $M+NH_4^+$ = 7265.74 and deconvoluted mass found 7250.88 and 7264.78.

Compound **11.3.13** (Beta conjugate with fluorescent peptides).

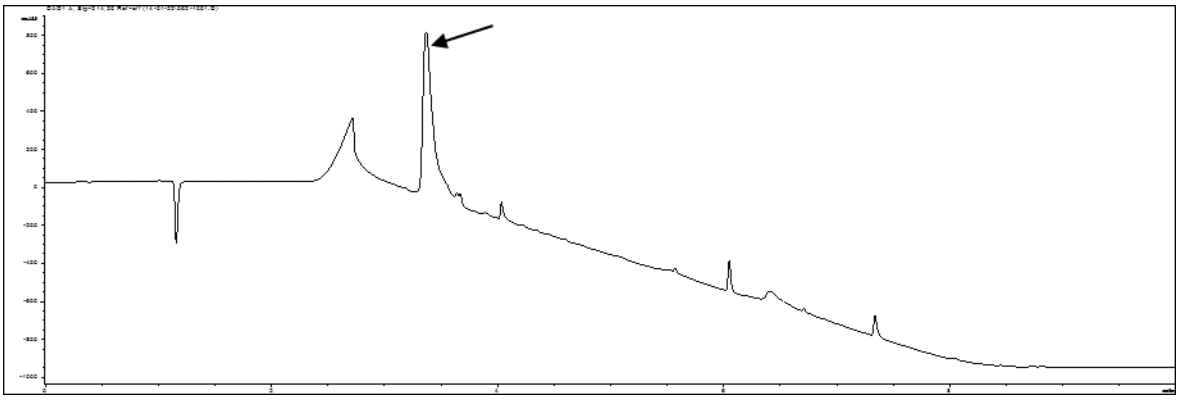


Figure 11.3.15: RP-HPLC trace of compound **11.3.13** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

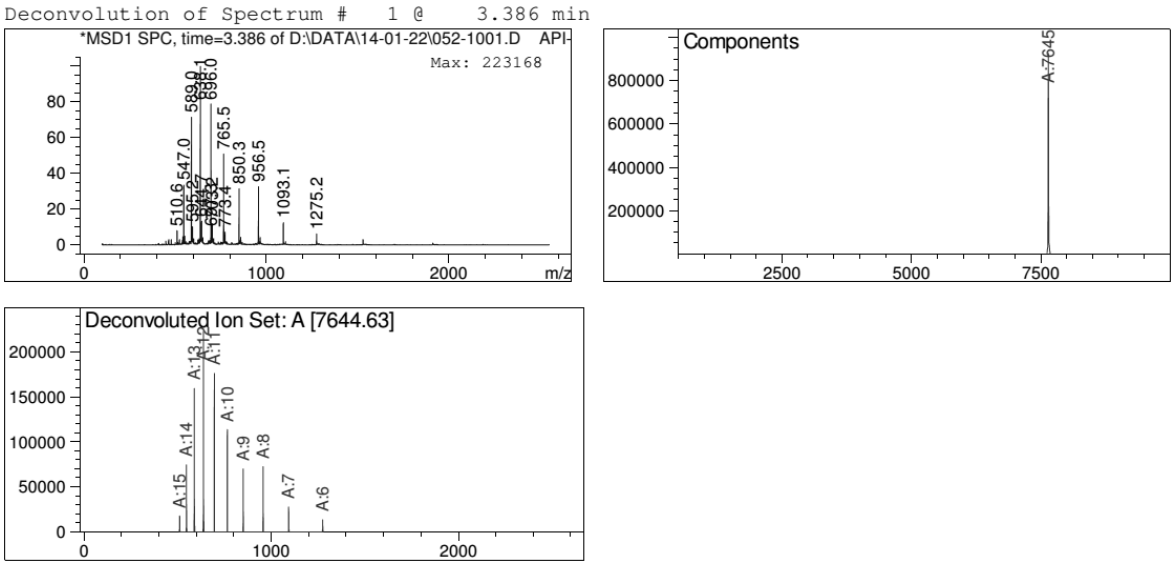


Figure 11.3.16: ESI-MS from LC-MS at t = 3.386 min. for pure compound **11.2.13**. E.M calcd. for $C_{322}H_{510}N_{104}O_{113}$ = 7641.74 and deconvoluted mass found 7644.63.

Compound **11.3.14** (Gamma conjugate).

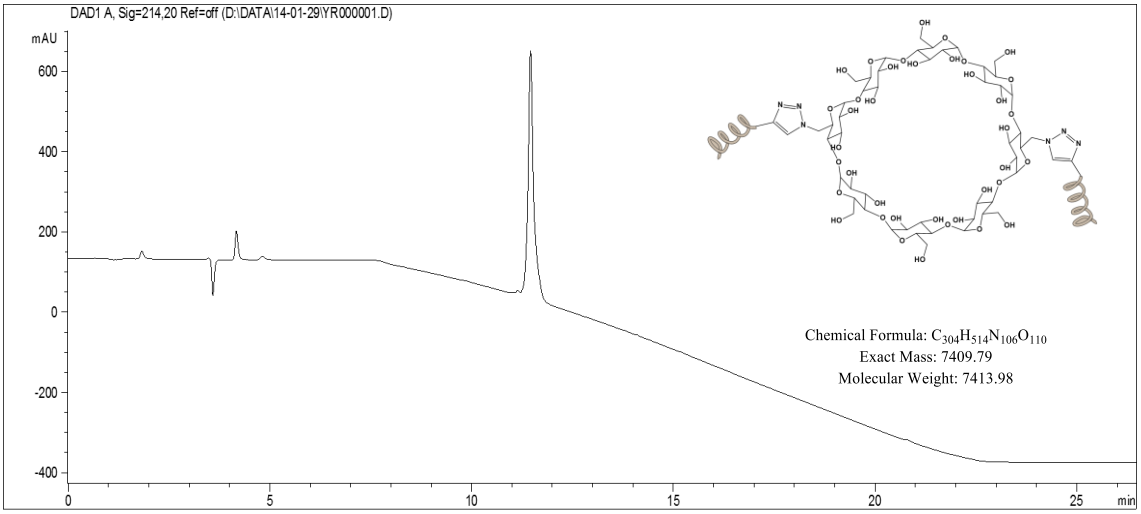


Figure 11.3.17: RP-HPLC Chromatogram of pure compound **11.3.14** (Jupiter C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

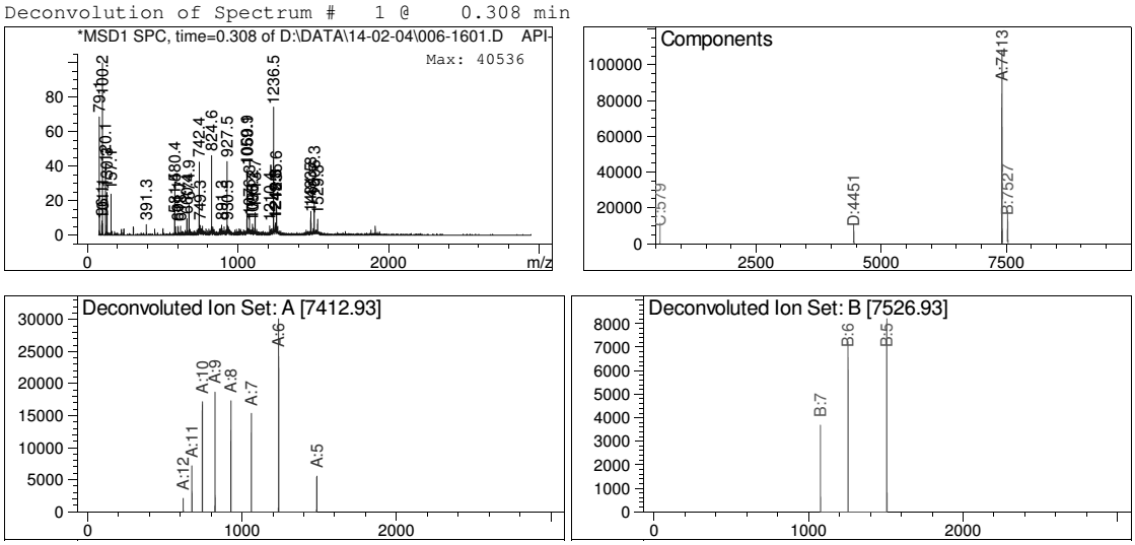


Figure 11.3.18: ESI-MS of compound **11.3.14**. Calcd. E.M. = 7409.79, M + 114.02 (TFA salt) = 7523.81 and deconvoluted masses found 7412.93 and 7526.93 + other minor impurities.

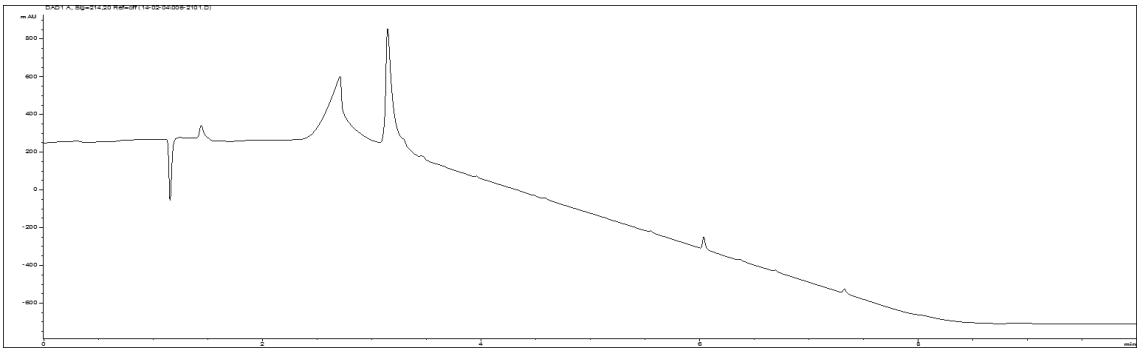


Figure 11.3.19: HPLC trace from LC-MS (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

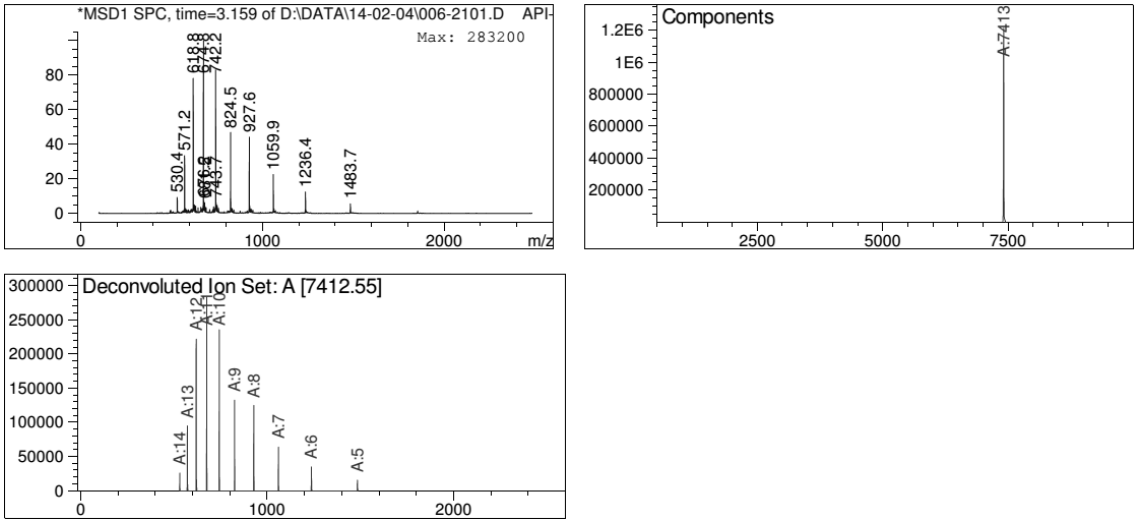


Figure 11.3.20: ESI-MS from LC-MS at r.t. = 3.159 min. Cacld. E.M. = 7409.79 and deconvoluted mass found 7412.55.

11.3.5 EMSA studies II

All gels have been run according to the protocols described in sections 10.4 & 10.5 (depending on whether non-radiolabelled or radiolabelled DNA was used):

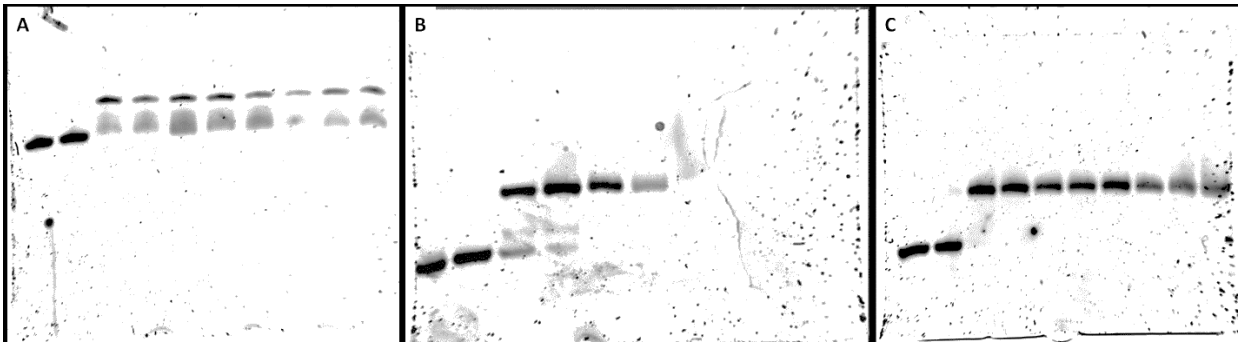


Figure 11.3.21: EMSA titration of the dipodal peptidocyclodextrin conjugates **11.3.11, 11.3.12** and **11.3.14** from top to botton to the CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3'). Lane 1-10: dsDNA (0.167 µM); lanes 1-10: 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of peptide.

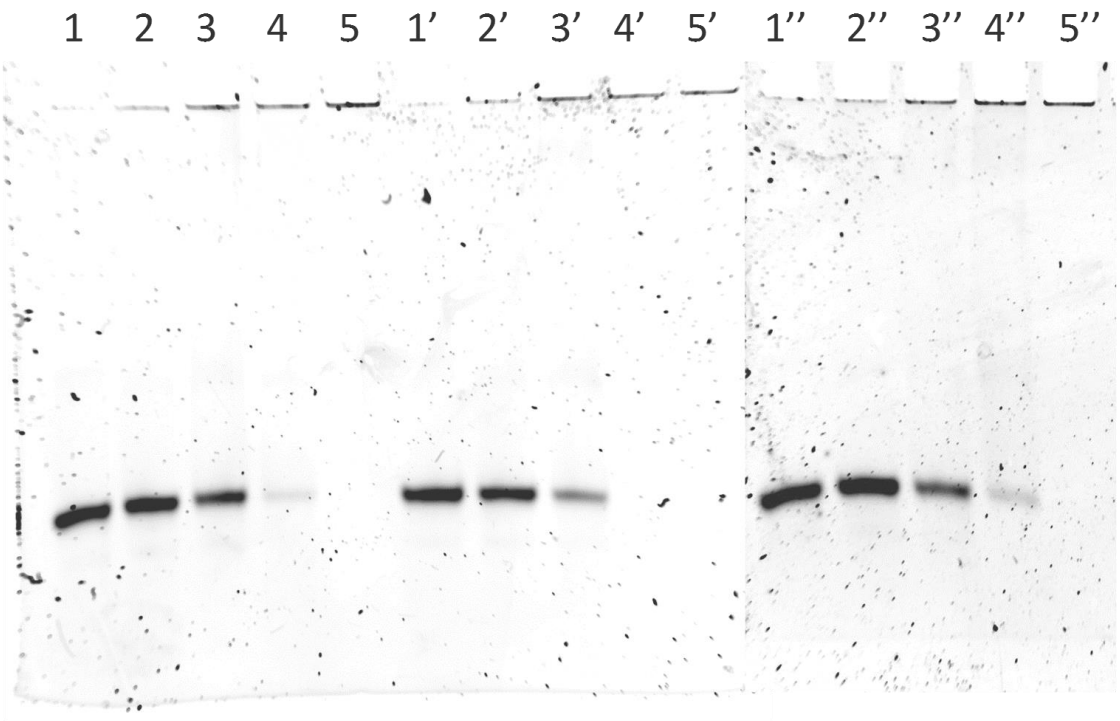


Figure 11.3.22: EMSA titration of the dipodal peptidocyclodextrin conjugates to the random sequence: Concentrations: 0.167 µM dsDNA; Lanes 1-5: 0. 0.167, 0.501, 0.668, 1.67 µM. Lanes 1-5 for compound **11.3.11** (alfa conjugate). Lanes 1’-5’ compound **11.3.12** (beta conjugate). Lanes 1’’-5’’ Compound **11.3.14** (gamma conjugate).

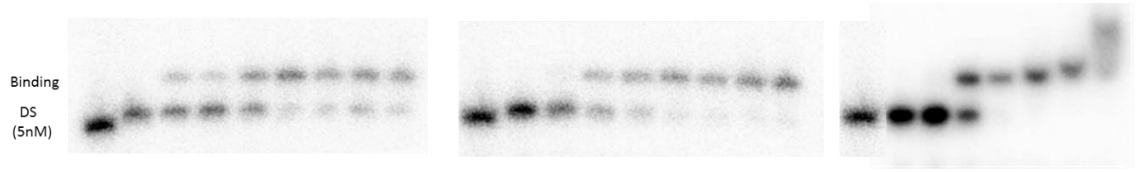


Figure 11.3.23: EMSA titration of the dipodal peptidocyclodextrin conjugates to the 5’-labeled ^{32}P-CRE sequence: Concentrations mentioned above. First lane in all the gels: pyrimidine strand. Gels from left to right: **11.3.11**, **11.3.12** and **11.3.14**.

CHAPTER 11.4

All cell lines, Fetal Bovine Serum, PBS (Gibco) and DMEM (Gibco) were provided by prof. Dr. Ir. B. De Geest. The synthesis of compounds relevant for this chapter have been described in chapters 11.2 and 11.3.

11.4.1 Confocal Microscopy: additional image

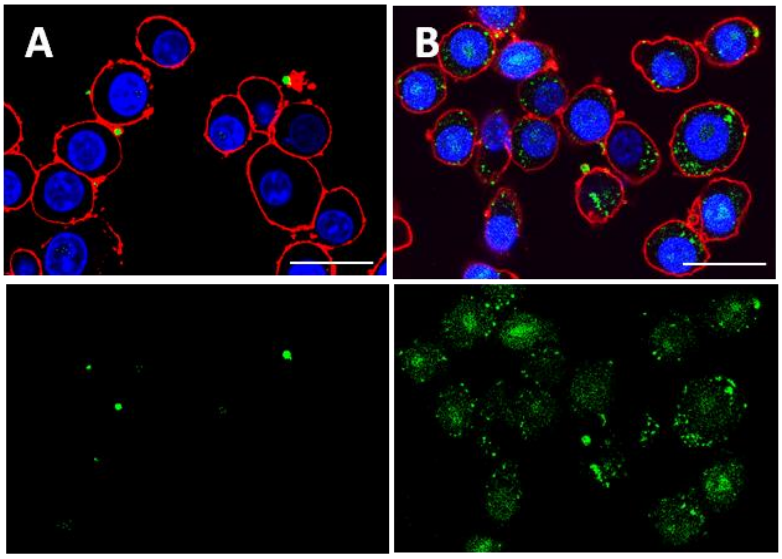


Figure 11.4.1: Confocal microscopy images of (A) the uptake of the GCN4 peptide **4.1** at 4 °C, (B) 37 °C Upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panel shows fluorescein image. (Scale bar = 20 µm).

From figure 11.4.1, it is clear that the GCN4 peptide (**4.1**) was internalized at 37 °C but not at 4°C pointing to an endocytic pathway.

11.4.2 Flow cytometry: additional information

The cell lines grow at a constant temperature of 37 °C in an 5 % CO_2 -atmosphere. A DMEM grow medium is enriched with fetal bovine serum, penicillin, pyruvate and glutamine after which the culture medium is sterilized through a filter. The medium, PBS and other reagents are always preheated at a temperature of 37 °C before they are added to the cells. To work under sterile conditions, cells are handled in a laminar flow cabinet (Nuaire).

Plating of cells

To perform the test on larger scale, 24- or 96-well plates are used. Therefore a specific amount of cells are divided along the wells. In order to divide the cells properly, the cells are counted in a Bürker counting chamber which is a glass microscope slide with a grid of perpendicular lines. For the macrophages (RAW264.7), the following protocol was used:

- Remove medium from the confluent flask using the vacuum system
- Wash the cells once with 10 ml of preheated PBS (37 °C)
- Add 5 ml of preheated culture medium
- Detach cells with the cell scraper
- Transfer cells to 15 ml tube and centrifuge (300 G for 5 min)
- Resuspend cells in culture medium
- From this tube, transfer 20 µl of cell suspension to an eppendorf and add 180 µl of trypan blue and mix well.
- Transfer the mixture to the Burker counting chamber to look at the cells with the microscope (Leica DMIL, NA 10x).

An average of the counts in 4 grids is multiplied by 10000 which gave the number of cells per mL. An amount of cells/well is chosen and a cell dilution is made based on the number of cells per mL. The cells are then plated in the well plates and incubated for 24 hrs at 37 °C and 5 % CO_2.

11.4.3 Serum stability assay

To a 100 µL solution containing 5 µM of Fluorescein labelled peptide dissolved in mQ, 10 µL Fetal Bovine Serum (FBS) was added and incubated at 37°C. 20 µL aliquots were taken after 0 h, 1 h, 2 h, 4 h and 24 h, diluted with 100 µL mQ and centrifuged for 30s. The supernatant was pipetted out and 100 µL was injected into the HPLC. The UV signal was measured at 490 nm. The % peptide remaining was calculated via manual integration by determining the area of the peak and comparing it to the area obtained for the same peak at t = 0 h.

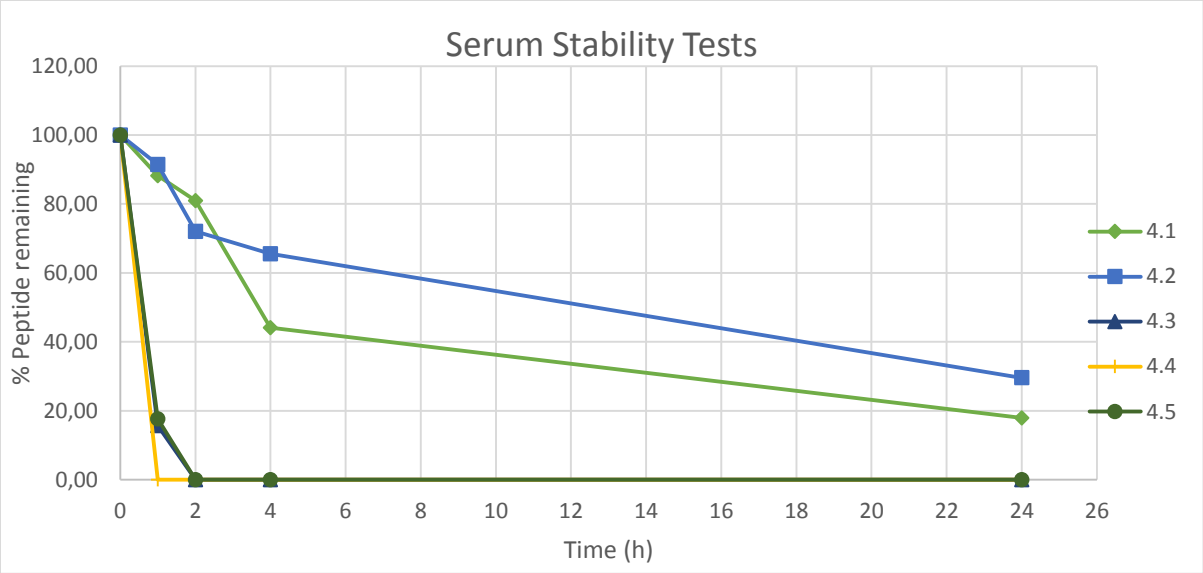


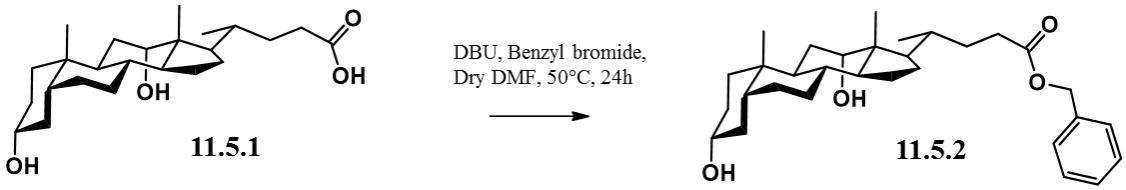
Figure 11.4.2: Graph of the serum stability assay by calculating the % peptide remaining using RP-HPLC via manual integration at $\lambda = 490$ nm.

The serum tests indicated that compounds **4.5**, **4.4** & **4.3** had degraded much faster than compounds **4.1** & **4.2**. The use of a cholic acid or a cyclodextrin scaffold in this case does not seem to enhance the stability of the peptides. Normally one would expect enhanced stability in serum due to the shielding effects possessed by these scaffolds, instead compounds **4.5**, **4.4** & **4.3** were found to degrade completely in 2 h or less. An explanation for this could be that the compounds **4.5**, **4.4** & **4.3** have a certain amount affinity for the serum proteins resulting in rapid degradation or they simply precipitate with the serum proteins when diluted and centrifuged. In case of **4.4** and **4.3** an alternate explanation could be the rapid cleavage of the ester bonds used for attaching the linkers to the scaffold. Due to the amphiphilic and complex nature of the compounds, it is however difficult to explain the unexpected, rapid degradation of the compounds in FBS via a simple serum test in vitro.

CHAPTER 11.5

11.5.1 Scaffold Synthesis

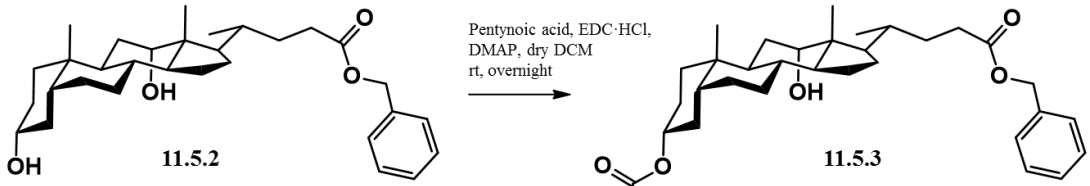
Benzyl 3*α*, 12*α*-dihydroxy-5β-cholan-24-oate (11.5.2). For synthesis, refer to **11.2.1**.



Chemical Formula: $C_{24}H_{40}O_4$
Exact Mass: 392,29
Molecular Weight: 392,57

Chemical Formula: $C_{31}H_{46}O_4$
Exact Mass: 482,34
Molecular Weight: 482,69

Benzyl 3*α*-pentynoate-5β-cholan-24-oate (11.5.3). Compound **11.5.2** (500 mg, 1.037 mmol) was dissolved in 2 mL dry DCM. Then a solution of 4-pentynoic acid (122.1 mg, 1.24 mmol) and DMAP (12.7 mg, 0.104 mmol) in dry DCM (2.2 mL) was added and stirred for 10 min in an ice bath. Then, EDC·HCl (298.3 mg, 1.55 mmol) in 1 mL dry DCM was added portion-wise to the reaction mixture and stirred for 30 min at 0°C, then the reaction mixture was stirred at room temperature overnight. The residue was purified by column chromatography (silica gel EtOAc/Hexane 1:3; R_f: 0.33) to give **11.5.3** as a yellow oil (349.9 mg, 60% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 5.15 (m, 2H), 4.75 (br, m, 1 H), 3.9 (m, 1H), 2.5 (s, 1H), 2.55-2.45 (m, 4 H), 2.1-2.0 (m, 2 H), 1.95-1.90 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 562.4 (100), found 621.3 (100) [M + OAc^-].



Chemical Formula: $C_{31}H_{46}O_4$
Exact Mass: 482,34
Molecular Weight: 482,69

Chemical Formula: $C_{36}H_{50}O_5$
Exact Mass: 562,37
Molecular Weight: 562,78

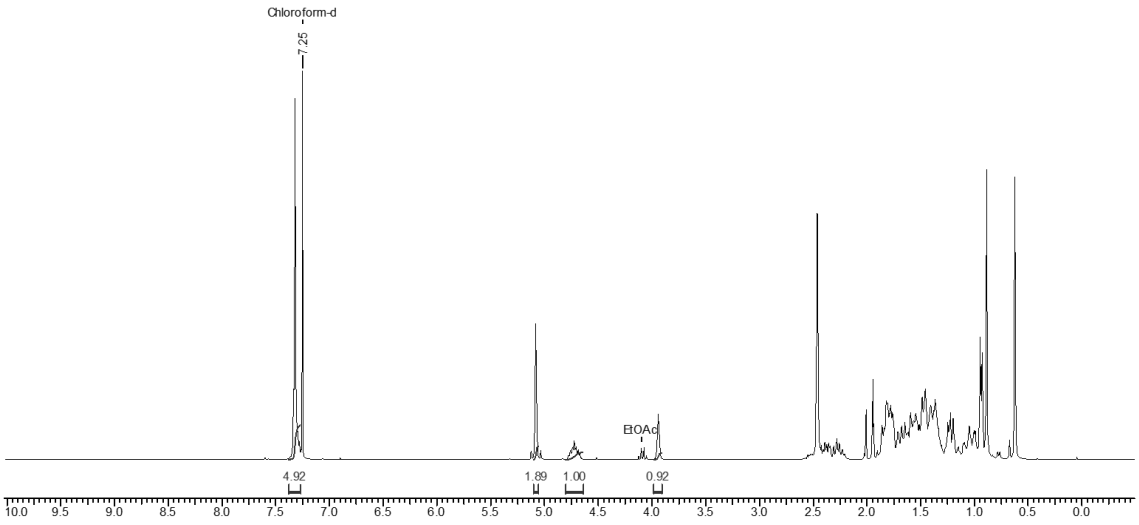
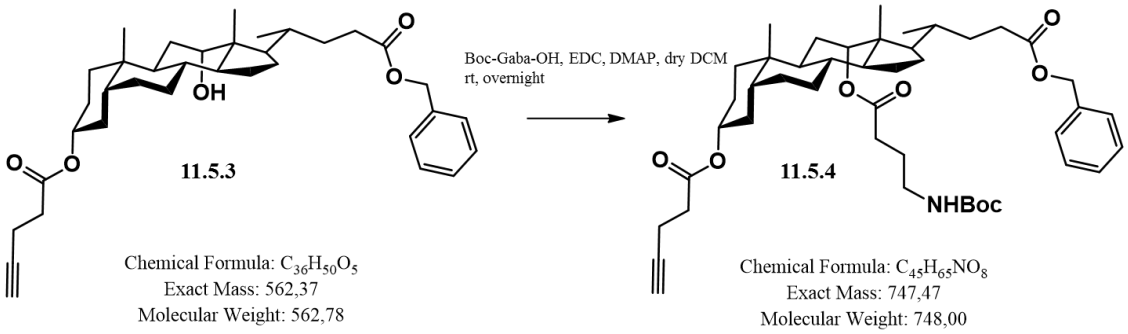


Figure 11.5.1. ^{1}H NMR spectrum of compound **11.5.3**.

Benzyl 3α-pentynoate-12α-Boc-aminobutyrate-5β-cholan-24-oate (11.5.4). Compound **11.5.3** (300 mg, 0.534 mmol) was dissolved in 1 mL dry DCM. Boc-GABA-OH (162 mg, 0.8 mmol) in 0.4 mL dry DCM and DMAP (130 mg, 1.068 mmol) in 0.4 mL dry DCM. The reaction mixture was allowed to stir at 0°C under argon for 10 min after which EDC was added portion wise and the stirring was allowed to continue for another 30 minutes in an ice bath. Finally, the reaction mixture was allowed to warm slowly and was left to stir overnight at room temperature. The residue was purified by column chromatography (silica gel Hexane/EtOAc 3:1; Rf: 0.38) to give **11.5.4** as a yellow oil (159.6 mg, 40% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 5.15 (m, 3H (12β + 2H)), 4.75 (br, m, 1H 3β), 3.15 (br m, 2H),2.60-2.50 (m, 2H), 2.55-2.45 (m, 4 H), 2.25 (br m, 2H), 2.1-2.0 (m, 2 H), 1.95-1.90 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 747.5 (100), found 806.5 (100) [M+ OAc]$^{-}$.



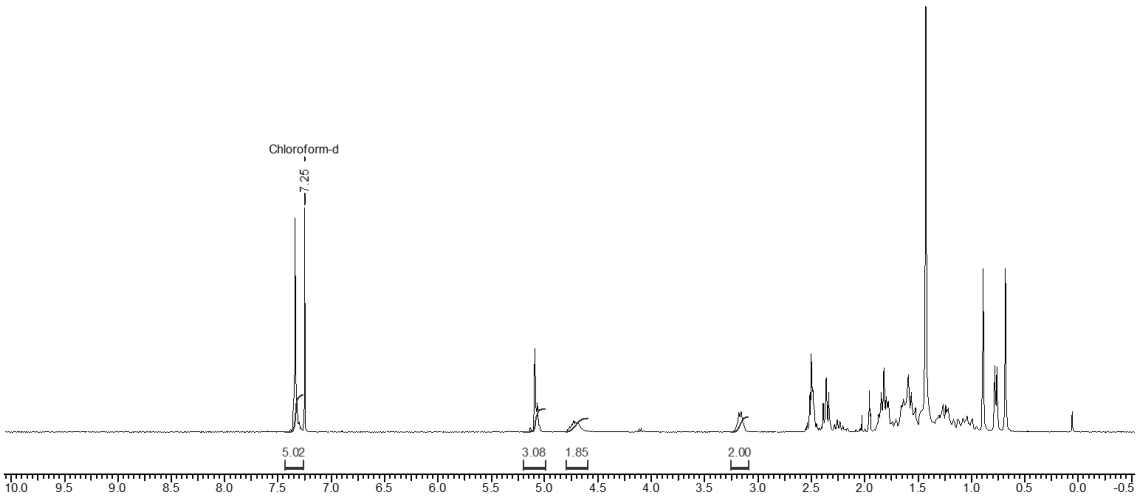
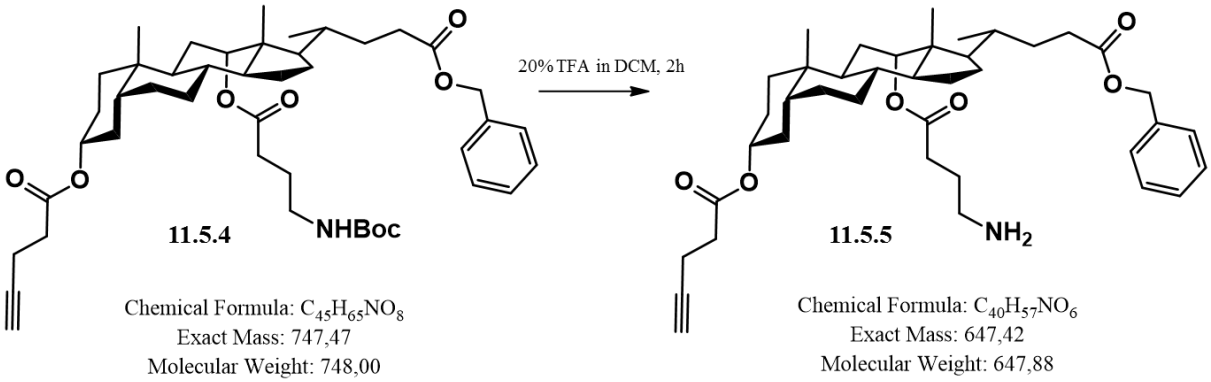


Figure 11.5.2. ^{1}H NMR spectrum of compound **11.5.4**.

Boc deprotection (11.5.5). Compound **11.5.4** (374.41 mg, 0.501 mmol) was dissolved in 4 mL DCM and 0.8 mL TFA was added. The reaction was stirred for 2h at room temperature. Then the solvent was evaporated and no further purification was needed (R_f: 0.13). Compound **11.5.5** obtained with 99% yield (321 mg). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 5.15 (m, 3H (12β + 2H)), 4.75 (br, m, 1H 3β), 3.15 (br m, 2H),2.60-2.50 (m, 2H), 2.55-2.45 (m, 4 H), 2.25 (br m, 2H), 2.1-2.0 (m, 2 H), 1.95-1.90 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ES-MS m/z (% rel. int.) calcd. 647.4 (100), found 648.4 (100) [M+ H^+].



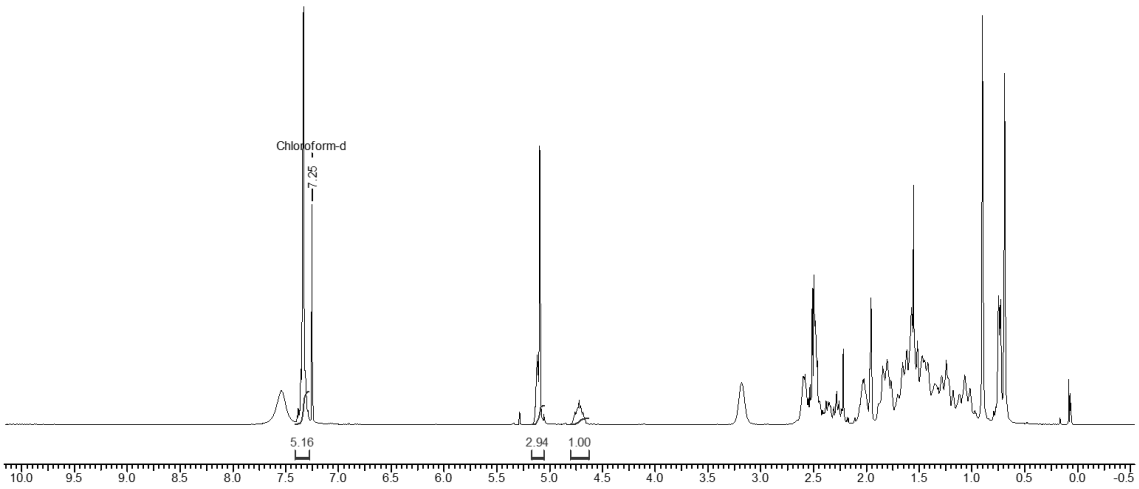
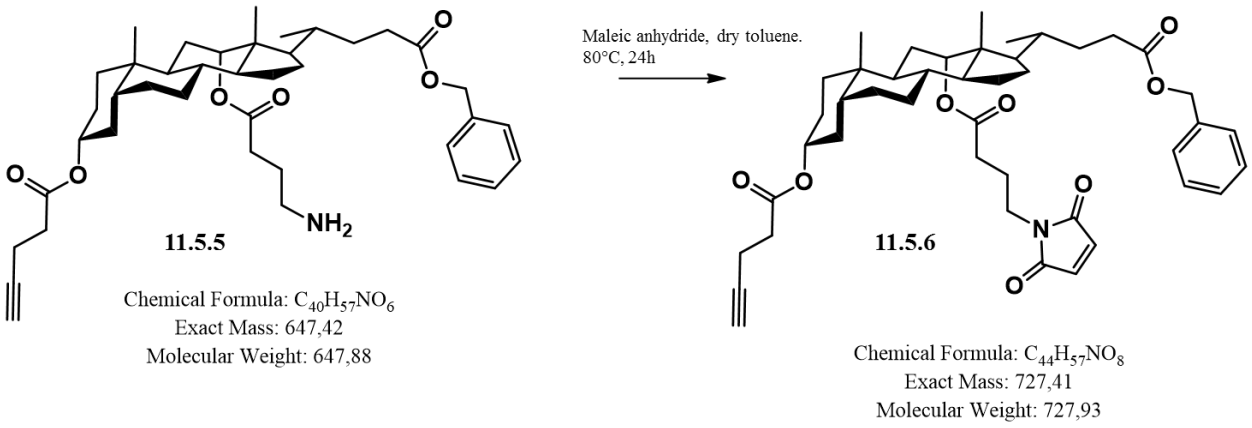


Figure 11.5.3. ^{1}H NMR spectrum of compound **11.5.5**.

Benzyl 3α-pentynoate-12α-4-maleimido-butyrate-5β-cholan-24-oate (11.5.6). Compound **11.5.5** (107.6 mg, 0.166 mmol) was dissolved in 2mL dry toluene. Maleic anhydride (32.56 mg, 0.332 mmol) is added to the reaction in 2 mL dry toluene. The reaction is refluxed at 80°C for 24h. The solvent was evaporated and the residue was purified by column chromatography (fine silica gel Hexane/EtOAc 4:1; Rf: 0.2) to give **11.5.6** as a brownish oil (45.9 mg, 38% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 6.65 (s, 2H), 5.17 (d, 1H, J = 13.57 Hz), 5.12 (d, 1H, J = 13.56 Hz), 4.75 (br, m, 1H 3β), 3.55 (t, 2H),2.55-2.50 (m, 2H), 2.45 (m, 4 H), 2.25 (br m, 2H), 1.95-1.90 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H).). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 174.13 (COOR-C24), 172.10 (2C maleimides), 171.46 (COOR-C3), 171.16 (COOR-C12), 136.88 (C - benzyl group), 134.51 (2CH maleimides), 128.88 (2CH-benzyl ring), 128.54 (2CH- benzyl ring), 128.20 (CH-benzyl ring), 76.34 (CH ester), 74.90 (CH ester), 69.00 (C-Alkyne), 66.33 (CH2-benzyl ester), 49.72 (CH), 47.93 (CH), 45.50 (C), 42.33 (CH), 37.56 (CH_2), 36.11 (CH), 35.16 (CH), 34.78 (CH), 34.47 (CH_2), 34.02 (C), 32.50 (CH_2), 32.20 (CH_2), 31.55 (CH_2), 31.20 (CH_2), 26.80 (CH_2), 26.65 (CH_2), 26,10 (CH_2), 25.05 (CH_2), 25.00 (CH_2), 24.51 (CH_2), 23.87 (CH_2), 23.21 (19-CH_3), 17.71 (21 – CH_3), 14.71 (CH_2), 12.54 (18-CH_3). ES-MS m/z (% rel. int.) calcd. 727.4 (100), found 745.4 (100) [M+ NH_4^+]. HR-MS (ES) m/z calcd. for $C_{44}H_{57}NO_8$ 727.4084, found 745.4448 (M + NH_4^+; Δ = 3.5 ppm).



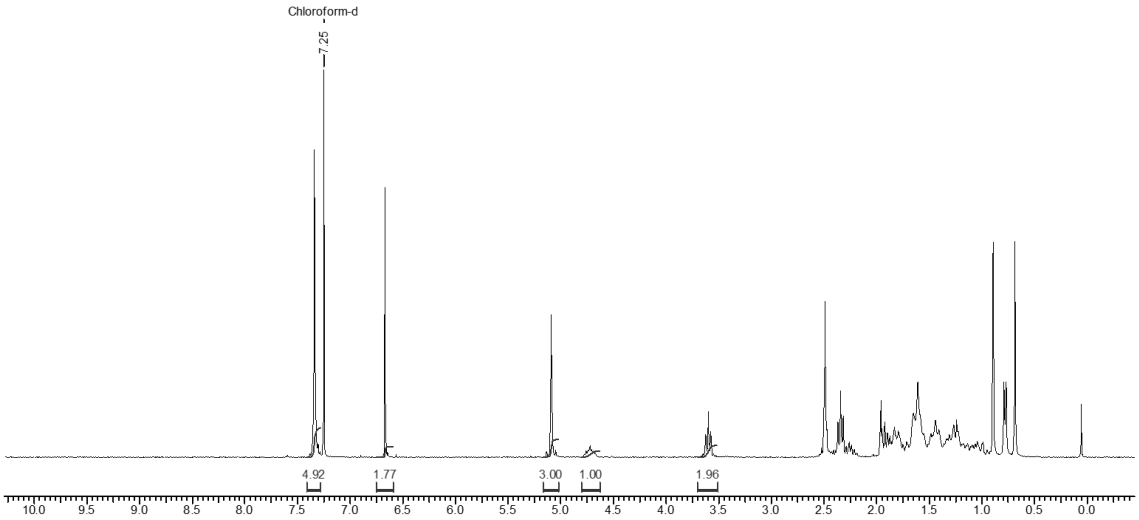


Figure 11.5.4. ^{1}H NMR spectrum of compound **11.5.6**.

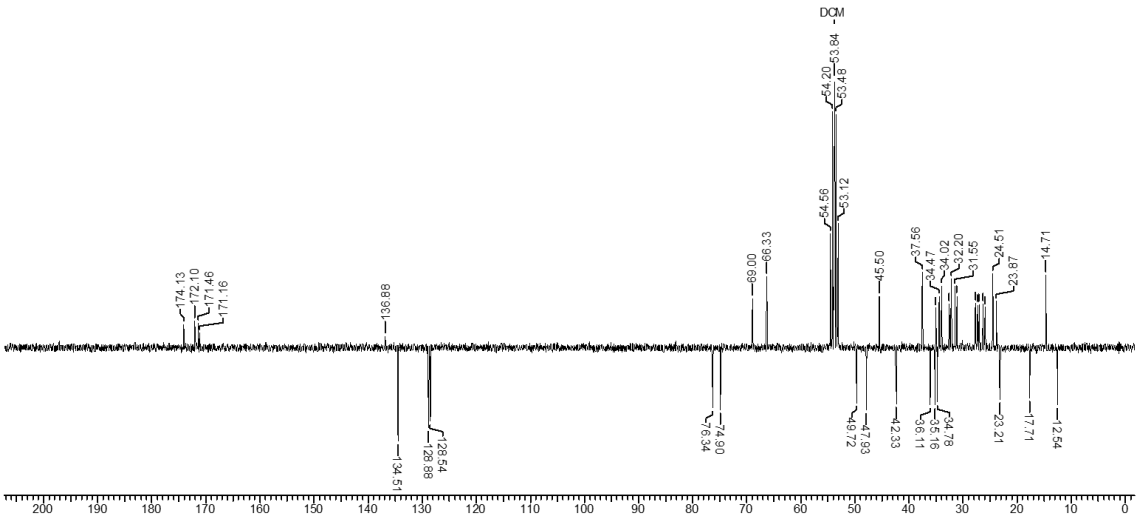
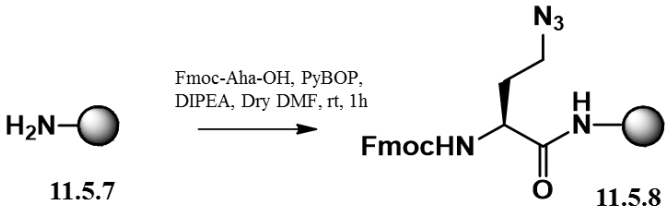


Figure 11.5.5. ^{13}C NMR spectrum of compound **11.5.6**.

11.5.2 Peptide synthesis:

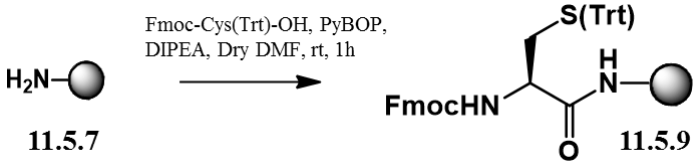
Manual coupling protocols

Immobilization of Fmoc-Aha-OH on Rink amide ChemMatrix resin



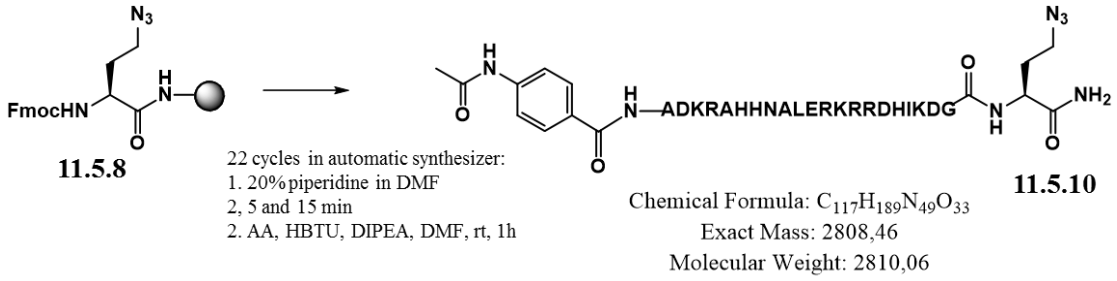
To a suspension of resin (200 mg, 0.54 mmol/g) in DMF (dry, 10 mL/g resin), were added Fmoc-Aha-OH (118.7 mg, 0.324 mmol), HBTU (122.87 mg, 0.324 mmol) and DIPEA (0.112 mL, 0.648 mmol). The mixture was shaken at room temperature for 1 h. The resin was then capped with acetic anhydride (0.061 mL, 0.648 mmol) and DIPEA (0.112 mL, 0.648 mmol) in dry DMF (3 mL) 2 times for 30 min.

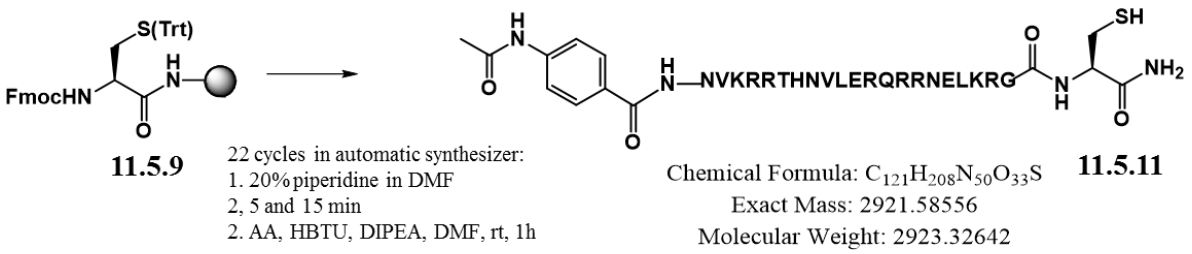
Immobilization of Fmoc-Cys(Trt)-OH on Rink amide ChemMatrix resin



At a suspension of resin (200 mg, 0.54 mmol/g) in DMF (dry, 10 mL/g resin), were added Fmoc-Cys(Trt)-OH (189.7 mg, 0.324 mmol), HBTU (122.9 mg, 0.324 mmol) and DIEA (0.112 mL, 0.648 mmol). The mixture was shaken at room temperature for 1 h. After the reaction, the resin was washed with DMF, ACN and MeOH. The resin was then capped with acetic anhydride (0.061 mL, 0.648 mmol) and DIPEA (0.112 mL, 0.648 mmol) in dry DMF (3 mL) 2 times for 30 min.

Automated synthesis of linear peptides





Automated solid phase peptide synthesis was carried out as described in 10.2.

The peptide was then cleaved for the resin and deprotected with a cocktail of TFA/TIS/water = 95: 2.5: 2.5 for 4h. After precipitation in cold ether, the peptide was analyzed by RP-HPLC and MALDI.

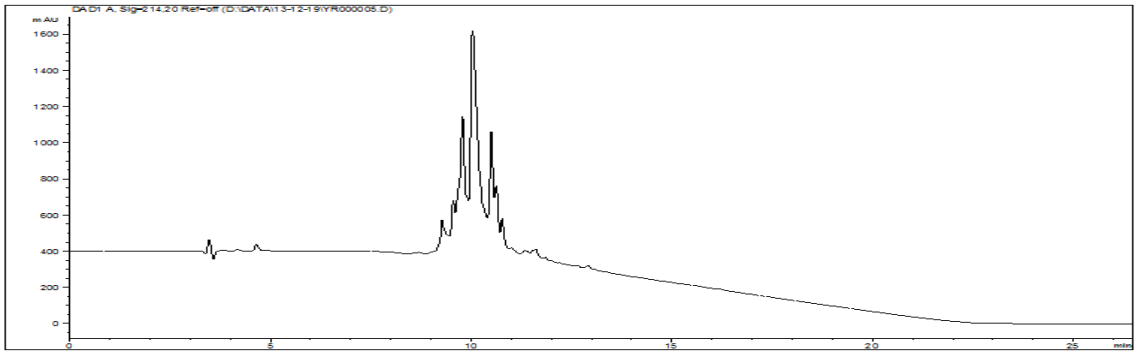


Figure 11.5.6: RP-HPLC Chromatogram of crude compound **11.5.10** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

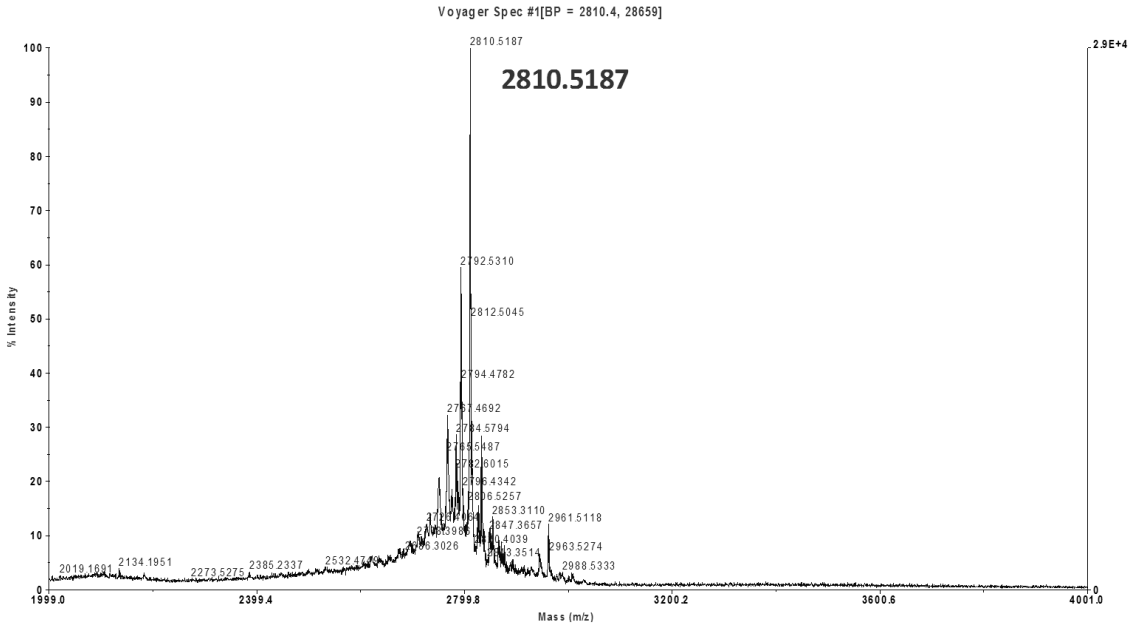


Figure 11.5.7: MALDI-TOF Spectrum of compound **11.5.10** from peak at t = 10.034 min. E.M. calcd. for $C_{117}H_{189}N_{49}O_{33}$ = 2808.46. Found M + H^+ = 2810.5187.

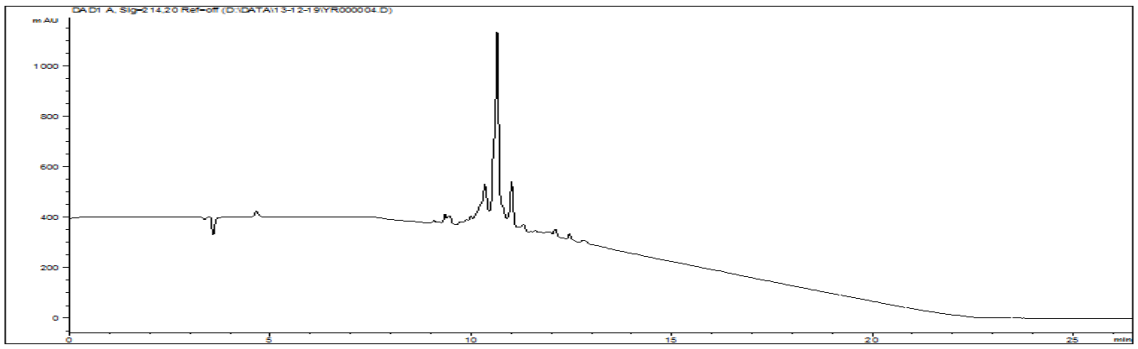


Figure 11.5.8: RP-HPLC chromatogram of crude compound **11.5.11** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

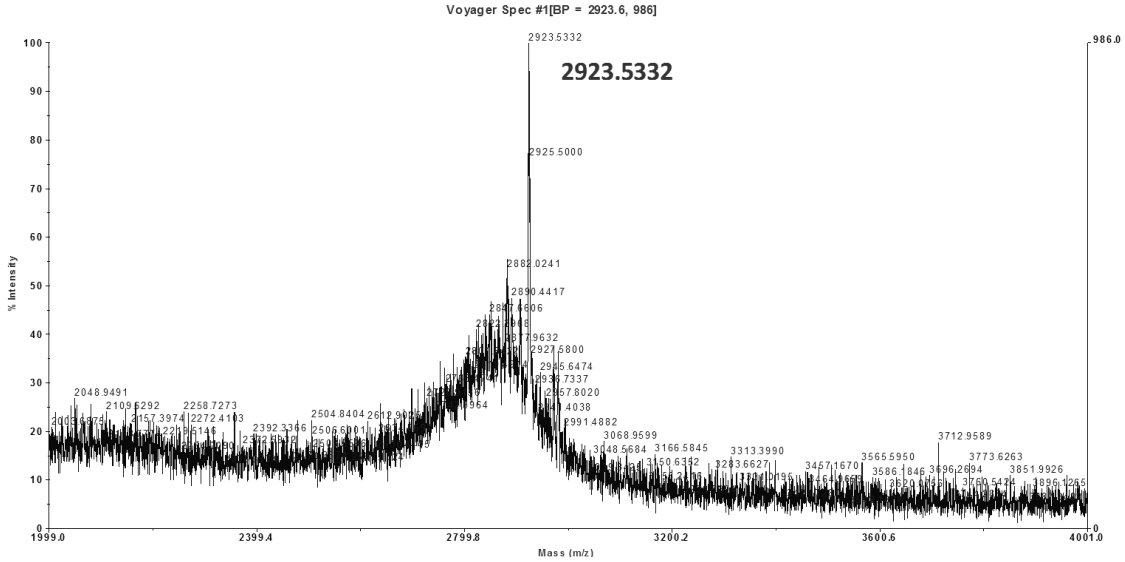
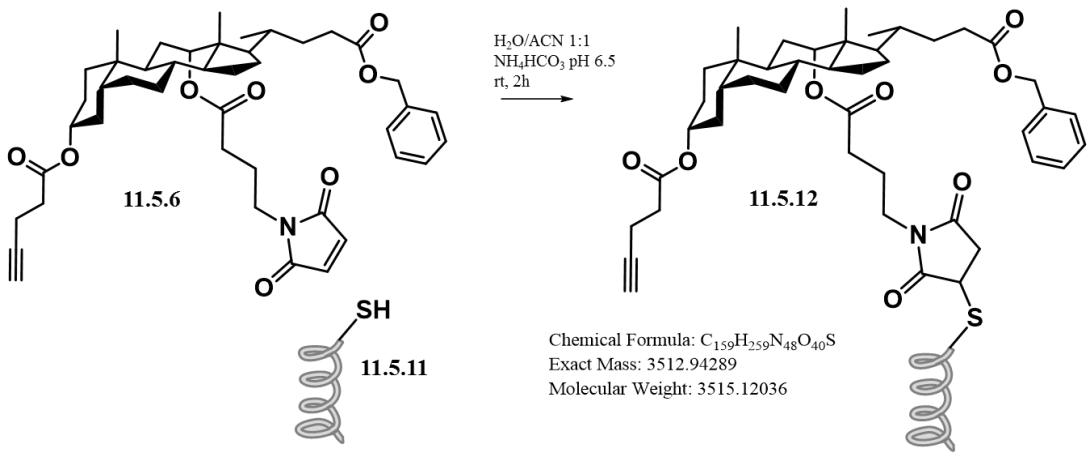


Figure 11.5.9: MALDI-TOF Spectrum of crude compound **11.5.11** from peak at t = 10.651 min. E.M. calcd for $C_{121}H_{208}N_{50}O_{33}S$ = 2921.59. Found: 2923.5332 = M + H^+.

11.5.3 Conjugation of peptides to the steroid scaffold

Conjugation of peptide cMyc to C12 position of scaffold

Scaffold **11.5.6** (1.24 mg, 0.002 mmol) and peptide **11.5.11** (8 mg, 0.003 mmol) were dissolved in 1 mL of a mixture H_2O/ACN (4:1). The pH was adjusted to 6.5 with NH_4HCO_3 buffer. The reaction was stirred for 2h and it was monitored by MALDI and HPLC. The solvent was evaporated and the crude compound **11.5.12** was subjected for CuAAC without prior purification.



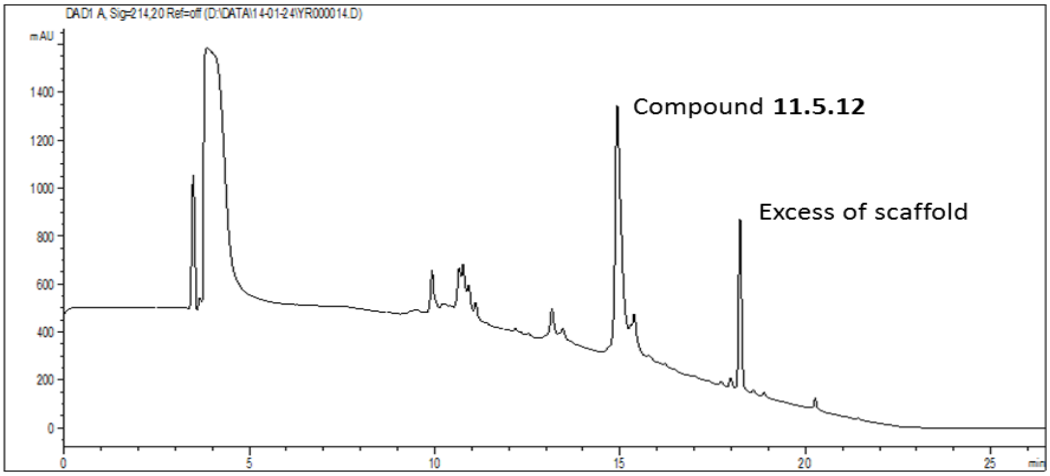
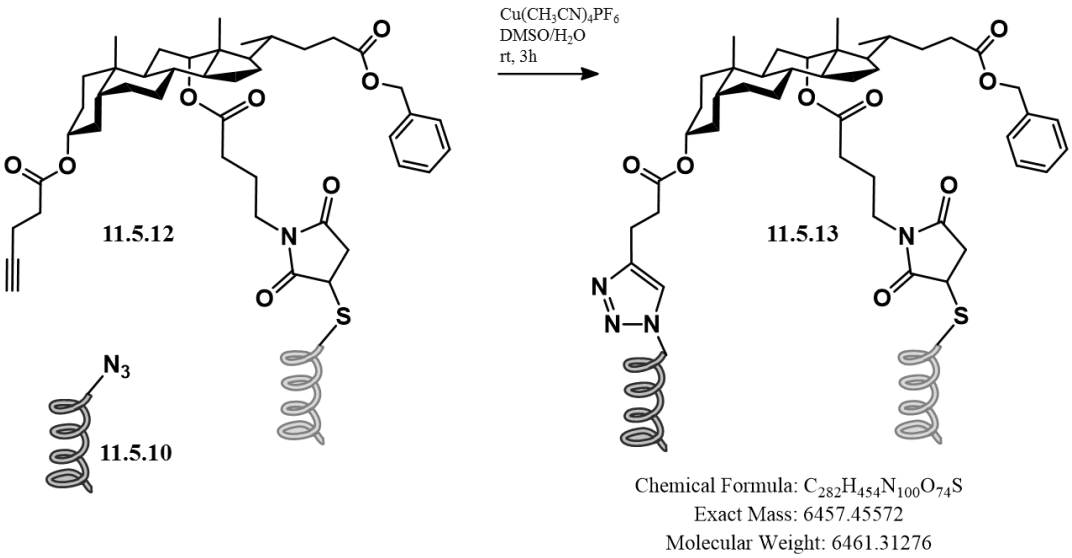


Figure 11.5.10: RP-HPLC Chromatogram of reaction mixture for synthesis of **11.5.12** after t = 2 h (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

Conjugation of peptide Max to C3 position of compound 11.5.12

Compound **11.5.12** (5 mg, 0.0013 mmol) was dissolved in 0.2 mL dry DMSO. Peptide **11.5.10** (6.15 mg, 0.0022 mmol) in 0.1 mL H_2O/ACN 4:1 was added to the scaffold and the reaction was stirred for 10 min under argon. $Cu(CH_3CN)_4PF_6$ (5.1 mg, 0.014 mmol) dissolved in 0.075 mL dry DMSO was added to the reaction mixture. The reaction was monitored by HPLC and it was completed after 3h. Compound **11.5.13** was purified by RP-HPLC.



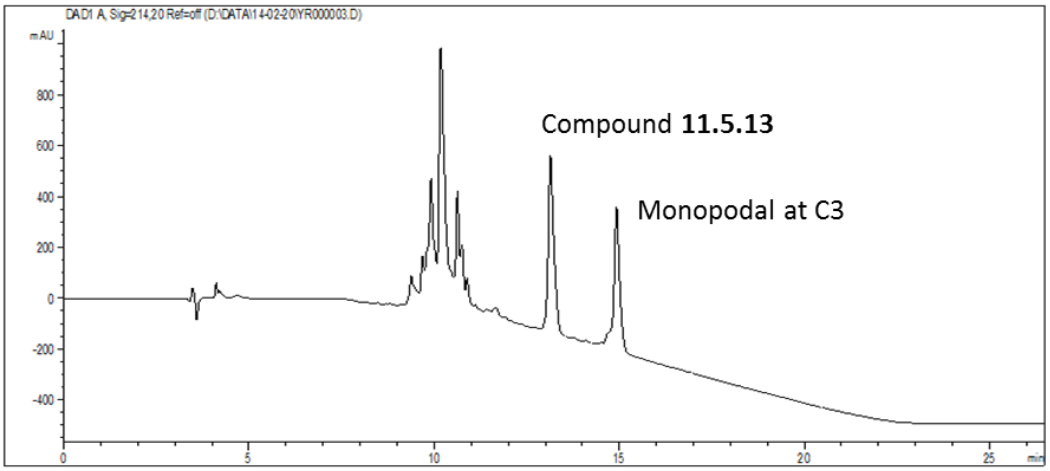


Figure 11.5.11: RP-HPLC Chromatogram of reaction mixture for synthesis of **11.5.13** at t = 3 h (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

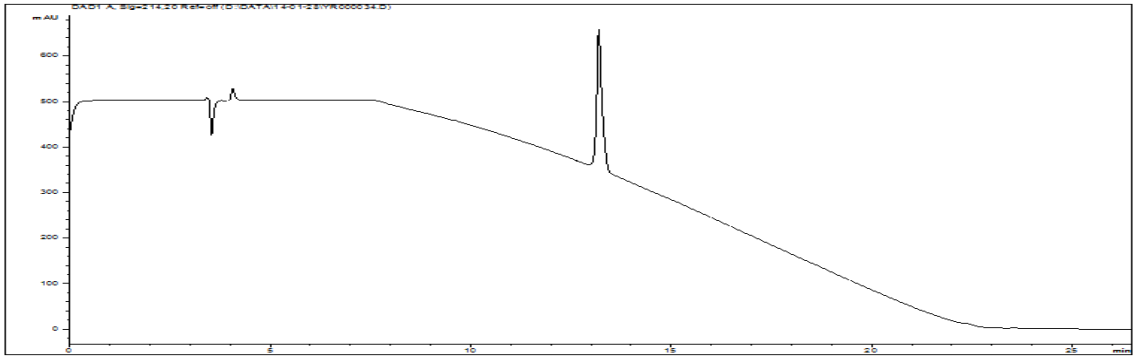


Figure 11.5.12: RP-HPLC Chromatogram of RP-HPLC purified compound **11.5.13** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

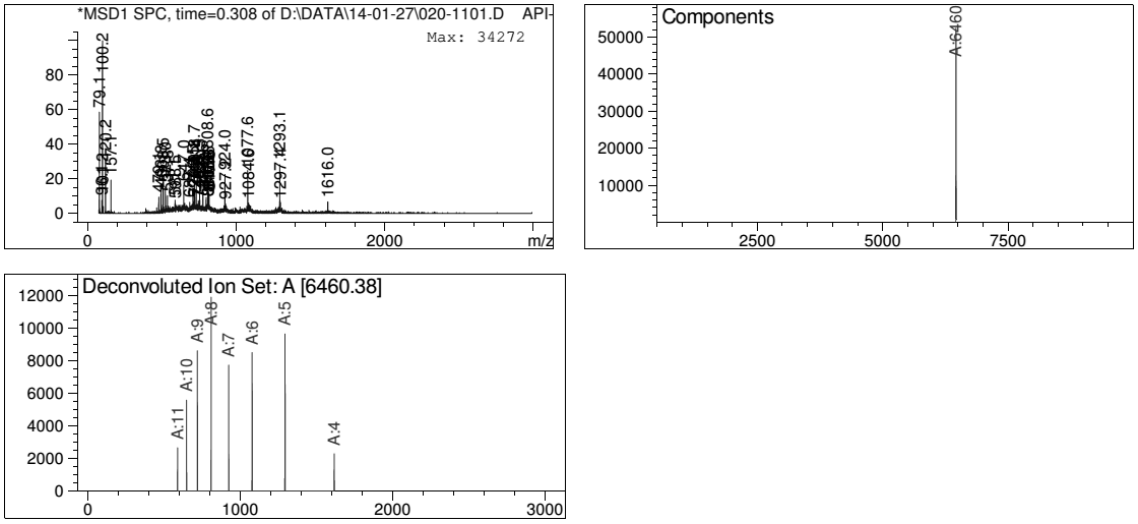


Figure 11.5.13: ESI-MS of RP-HPLC purified compound **11.5.13**. E.M. calcd. for $C_{282}H_{454}N_{100}O_{74}S$ = 6457.46 and deconvoluted mass found 6460.38.

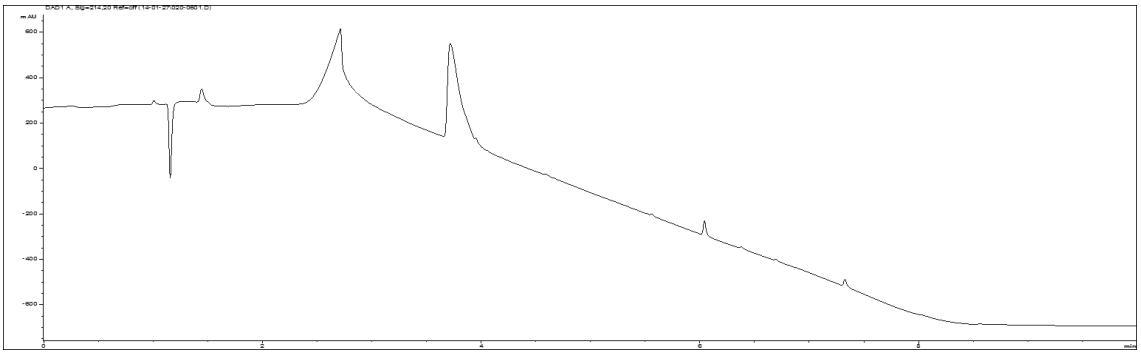


Figure 11.5.14: RP-HPLC trace from LC-MS for HPLC purified compound **11.5.13** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C).

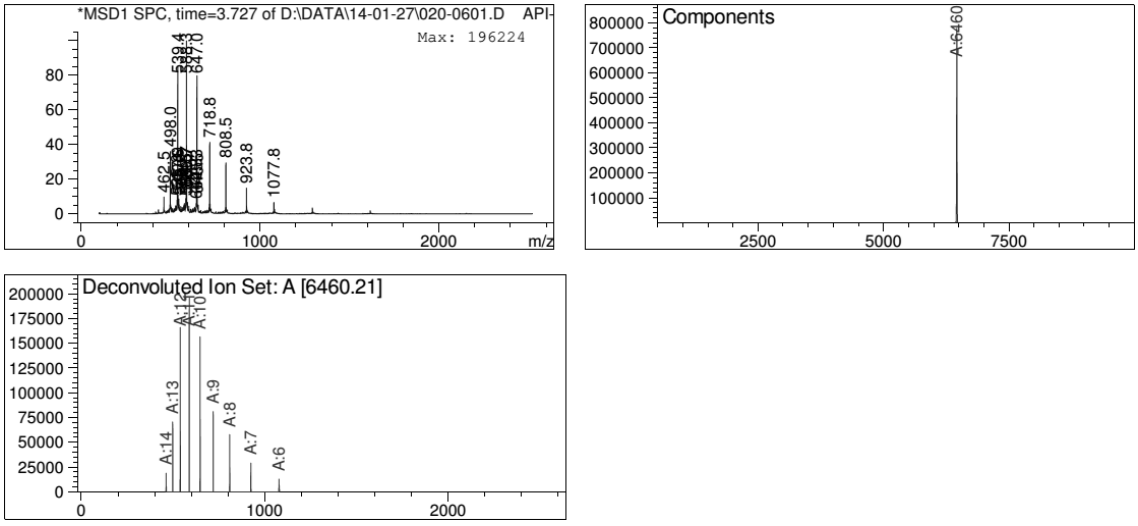


Figure 11.5.15: ESI-MS from LC-MS at r.t. = 3.727 min. E.M. calcd. for $C_{282}H_{454}N_{100}O_{74}S$ = 6457.46 and deconvoluted mass found 6460.21.

11.5.4 EMSA studies III

All gels have been run according to the protocols described in sections 10.4 & 10.5 (depending on whether non-radiolabelled or radiolabelled DNA was used):

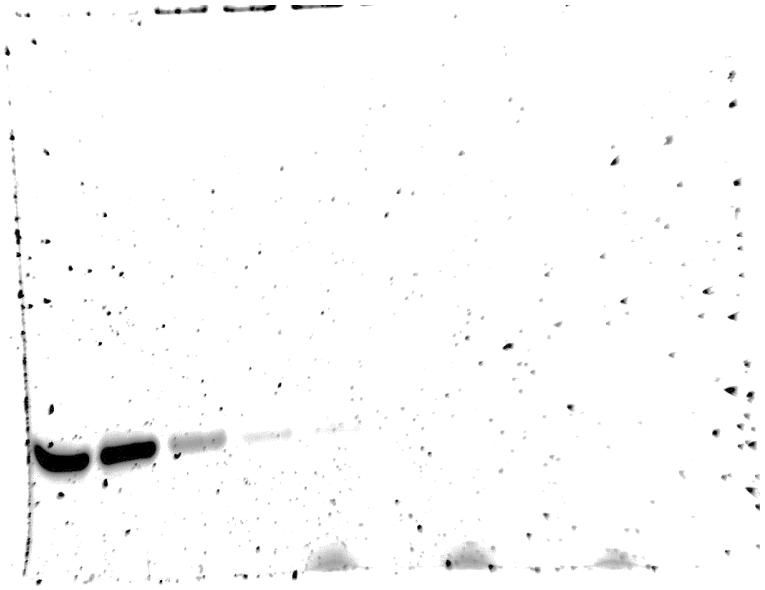


Figure 11.5.16: EMSA titration of the dipodal peptidosteroid **11.5.13** to the ramdom dsDNA stained with SybrGold.

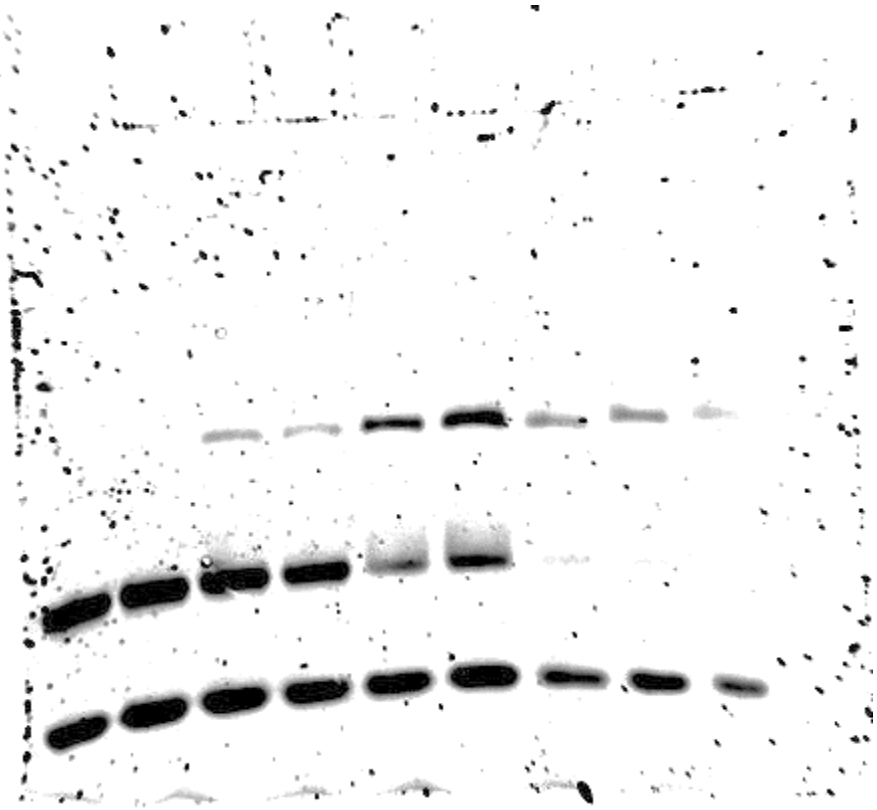


Figure 11.5.17. EMSA titration of the dipodal peptidosteroid **11.5.13** to the E-box binding site stained with SybrGold.

CHAPTER 11.6

11.6.1 Chapter Specific Material and Methods

Special instructions

All manual reactions involving the photocleavable linker were shielded from direct light exposure by foil wrapping. These means were complemented by standard laboratory glassware. Reactions were usually performed at room temperature. Occasional cooling was established by using a special double-walled glass reactor (~7 mL), combined with a Julabo F70 cryostat controlled by a Julabo VC. Photolyses were carried out on a small scale with a 4W Bioblock Scientific compact UV lamp, set at 365 nm. For large scale cleavages, a 451 W UV ACE glass incorporated 7225-34 immersion lamp equipped with a Schott WG320 UV cut-off filter was used. The resin suspension was thereby continuously agitated by N_2 bubbling, and solvent regularly replenished, to prevent direct resin irradiation by complete solvent evaporation. All samples were cleaved at a distance of ~1 cm from the lamp.

Analytical techniques and instrumentation

Occasionally measured resin loadings were calculated from UV-data collected with a Varian Cary 3E UV-VIS spectrophotometer at room temperature, operating in double beam mode, using standard quartz cuvettes (l = 1 cm, ~600 µL).

By using a binary solvent system composed of (A =) H_2O + TFA (0.1 %) and (B =) ACN as mobile phases, linear gradient elution has been performed. After injection, the column was flushed with **x %** B for 3 min, followed by a linear increase of B (versus A) to 100 % in **y** min, finishing by flushing with 100 % B for 5 min, after which the gradient returns to **x %** B in 0.5 min, concluding the cycle by flushing with **x %** B for 3 min.

- **Gradient 1**: refers therein to (**x**,**y**) = (0, 15), or a 0 to 100 % linear increase of B in 15 min
- **Gradient 2**: refers therein to (**x**,**y**) = (75, 15), or a 75 to 100 % linear increase of B in 15 min

In addition to these, all other gradients follow the same profile.

All MALDI spectra were recorded in the positive and reflector mode, with delayed extraction. 2,5-dihydroxybenzoic acid (DHB) was 98.0 % pure, and α-cyano-4-hydroxycinnamic acid (α-CHCA) was 99 % pure. All spectra were calibrated against MePEOH ($M_n \approx 2000$, PD = 1.06), spotted from a MeOH (2 mg/mL) solution. Following matrix solutions were utilized (made in eppendorf tubes, stored in freezer, carefully defrosted and homogenized upon use):

Predominantly used (and referred to in presented analytical data):

- **Standard DHB**: DHB (10 mg) + ACN (500 µL) + H_2O (470 µL) + TFA_{aq} (30 µL, 3 %)
- **Standard α-CHCA**: α-CHCA (10 mg) + ACN (500 µL) + H_2O (400 µL) + TFA_{aq} (100 µL, 3 %)

Occasionally tested (alongside Table 4.10, Figure 4.5 and Figure 5.5):

- Alternative DHB 1: DHB (10 mg) + ACN (500 µL) + MeOH (470 µL) + TFA_{aq} (30 µL, 3 %)
- Alternative DHB 2: DHB (1 .54 mg) + $CHCl_3$ (666 µL) + MeOH (333 µL)

By using a binary solvent system composed of (A =) H_2O + formic acid (0.1 %) or H_2O + NH_4OAc (5 mM) and (B =) ACN as mobile phase, linear gradient elution has been performed. After injection: flushing the column with **x %** B for 2 min, followed by a linear increase of B (versus A) to 100 % in 15 min, finishing by flushing with 100 % B for 5 min, after which the gradient returns to **x %** B in 0.5 min, concluding the cycle by flushing with **x %** B for 3 min.

Similar to aforementioned RPHPLC profiles, yet with a shorter (2 instead of 3 min) initial flushing phase and different additives (formic acid or NH_4OAc instead of TFA) the following gradients have also been used:

- **Gradient 1**: refers therein to **x** = 0, or a 0 to 100 % linear increase of B (versus A) in 15 min
- **Gradient 2**: refers therein to **x** = 75, or a 75 to 100 % linear increase of B (versus A) in 15 min

Recurring SPPS procedures

- **Colour testing**

TNBS test[1]

A minute resin sample (~1-5 mg) was transferred to a miniature glass test tube, to which a DIPEA solution in DMF (10 % v/v, 4 drops) and a TNBS solution in DMF (2 % v/v, 4 drops) were added. Intense, orange/red coloration of beads indicates presence of free primary amine functions. Unaffected beads support absence (i.e. derivatization) of these functions, yet results are only indicative. Comparison with starting material or reference compounds has generally been made for reliable interpretation. Since sensitive to steric impediment, prolonged (> 30 min) treatment might be required or test might fail for hindered substrates, while premature NHFmoc deprotection by the DIPEA component might bias conclusions.

NF31 test[2]

A minute resin sample (~1-5 mg) was transferred to a mini glass test tube, to which an NF31 solution in ACN (2 mM, 4 drops) was added. The mixture was heated for 10 minutes at 70 °C. The reagent solution was carefully removed and the resin repeatedly washed with small volumes of DMF, MeOH, ACN and

[1] Provides for detection of primary amines .

Hancock, W. S.; Battersby, J. E. *Anal. Biochem.* **1976**, *71*, 260.264. A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzenesulphonic acid.

[2] Provides for detection of both primary, secondary and aromatic amines.

(a) Madder, A.; Farcy, N.; Hosten, N. G. C., De Muynck, H.; De Clercq, P. J.; Barry, J.; Davis, A. P. *Eur. J. Org. Chem.* **1999**, 2787-2791. A novel sensitive colorimetric assay for visual detection of solid-phase bound amines.

(b) Van der Plas, S. E.; De Clercq, P. J.; Madder, A. *Tetrahedron Lett.* **2007**, *48*, 2587-2589. Fast and easy detection of aromatic amines on solid support.

DCM. Being less affected by steric impediment and claimed more sensitive, red-pink coloured beads indicate presence of free amino functions (both primary, secondary and aniline-like), while completely colourless beads are aimed for confirming adequate derivatization.

- **UV-Fmoc loading determination**

A carefully weighed amount of thoroughly dried resin material is transferred to a volumetric flask, to which DCM (400 µL) and piperidine (400 µL) is added, aiming to ensure adequate homogenization at the bottom of the flask, preventing sticking of beads to the neck.[1] For 30 minutes, suspension was occasionally swirled. MeOH is added (1.6 mL) and DCM completes the exact flask volume, after which the mixture is left to settle. A blank solution, lacking the resin, is similarly prepared. Aiming for ~0.1 mM concentration to ensure linearity of the Lambert-Beer law, the appropriate flask volume and amount of resin are estimated via theoretical, calculated loadings, prior to the experimental procedure.

An aliquot of resulting solution is transferred to a standard quartz cuvette (l = 1 cm), and UV-absorption of the generated piperidine-dibenzofulvene adduct is measured at both 300 and 301 nm. At 300 nm, the concentration was calculated from a previously-established calibration curve,[2] while the reported ε_{301} = 7800 $M^{-1} \cdot cm^{-1}$ allowed for alternative determination via the Lambert-Beer law: Abs = $\varepsilon \cdot c \cdot l$. Taking flask volume and weighed amount of resin into account, experimental loadings could be estimated, usually averaging values.

- **Small scale release of photolinker-immobilized compounds**

Either aiming analytical verification or proceeding towards compound isolation, a resin sample (~1-5 mg) is transferred to a miniature glass test tube and suspended in the appropriate solvent (~4-8 drops), which has usually been ACN or EtOH. The tube was flushed with argon and sealed with a septum. Placed near-horizontally at a distance of ~1 cm from the small scale UV-lamp (365 nm), resin was irradiated for ~3 hours, with occasional manual homogenization of the suspension. Using a glass syringe with a narrow bore, cemented needle, resulting solution was carefully separated from the resin, ready for further manipulation or analysis.

- **Reagent B acidolytic cocktail** [3]

Consists of 88 % TFA (8.8 mL), mixed with 5 % H_2O (500 µL), 2 % TIS (200 µL) and 5 % phenol (weighed, 535.5 mg, ~500 µL) prior to addition. Maroon-purple coloration indicates decomposition after prolonged storage.

---

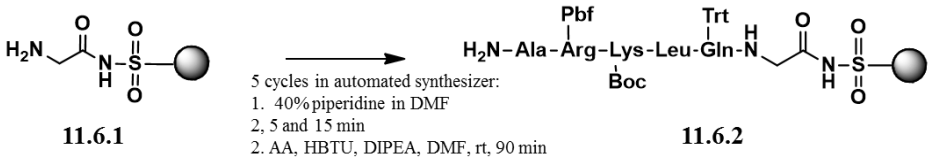
1 Adapted from The Combinatorial Index (Bunin, B. A.; *Academic Press* **1998**). And: ww.combinatorial.com

2 Nadia Farcy, PhD thesis, **2002**: Efforts towards the synthesis of a tripodal library for the discovery of serine-protease mimics.

3 Solé, N. A.; Barany, G. *J. Org. Chem.* **1992**, *57*, 5399-5403. Optimization of solid-phase synthesis of [Ala8]-dynorphin A.

11.6.2 Synthesis of donor peptides on Ellman's Safety-Catch resin

SPPS elongation of peptide building blocks on Ellman's loaded donor resin

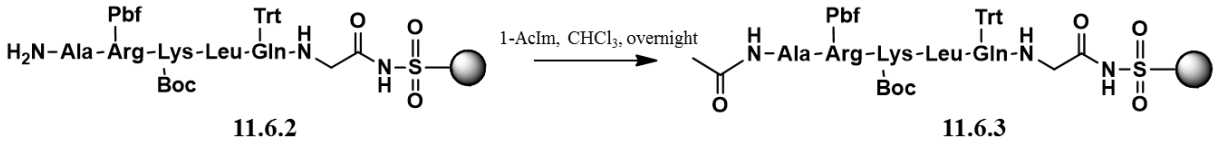


11.6.1 **11.6.2**

Starting resin **11.6.1** were transferred to SYRO-tubes of appropriate size.

Equal volumes of AA/DMF (0.5 M stock solution), DIPEA (dry)/NMP (0.9 M stock solution) and HBTU/DMF (0.5 M stock solution) were sequentially added to the SYRO reactor(s), ultimately supplying 5 equivalents of active species to the resin(s). Single couplings were performed in 1 ½ hour, with intermediate vortexing. DMF was used as washing solvent and a piperidine/DMF (40 % v/v) solution was applied for Fmoc-deprotection for 2, 5 and 15 min.

Capping of the N-terminus with 1-acetylimidazole

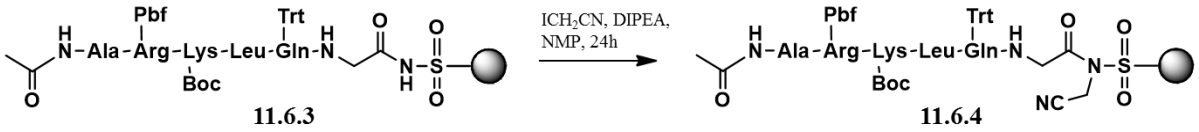


11.6.2 **11.6.3**

Resin **11.6.2** was manually washed with ACN, DCM and $CHCl_3$, prior to overnight capping of the N-terminus with 1-AcIm (10 eq) in $CHCl_3$ at room temperature to obtain compound **11.6.3**.

TNBS test: colourless

Activation by cyanomethylation of peptide-bearing Ellman's 4-sulfamylbutyryl safety-catch donor resin

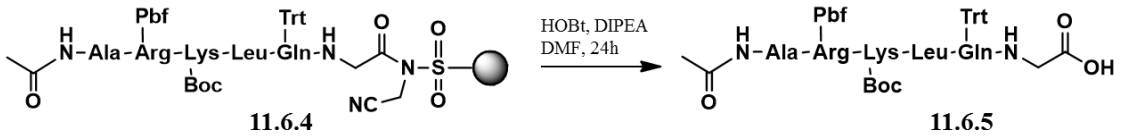


11.6.3 **11.6.4**

Beads **11.6.3** (0.184 mmol/g, 0.096 mmol) where washed with NMP prior to cyanomethylation. In the original SYRO tube (10 mL), swollen resin was resuspended in NMP (1.5 mL). Dry DIPEA (85 µL, 0.48 mmol, 5 eq.) and ICH_2CN (140 µL, 1.92 mmol, 20 eq.) were added and the pale yellow-orange mixture flushed with argon. Reaction proceeded for 24 hours at room temperature.

Significantly darkened, brown mixture was filtered under reduced pressure, resin **11.6.4** was washed with DMF, ACN, DCM and Et_2O and thoroughly dried under high vacuum.

Peptide cleavage & analysis



11.6.4 **11.6.5**

Chemical Formula: $C_{67}H_{93}N_{11}O_{14}S$
Exact Mass: 1307,66
Molecular Weight: 1308,59

In order to check if the synthesis of the peptide was complete, a small portion of resin **11.6.4** was cleaved (20 mg, 0.0037 mmol, 0,185 mmol/g), using HOBt (5 mg, 0.037 mmol, 10 eq) and DIPEA (6.5 µL, 0.037 mmol, 10 eq) were added to the resin suspended on dry DMF (160 µL). The reactor was flushed with argon and stirred for 24 hours at room temperature.

The supernadant solution containing the peptide **11.6.5** was filtered under reduced pressure and the solvent was evaporated.

Side-chain protected, C-terminally hydrolyzed AcHN-ARKLQG-OH **(11.6.5)** $C_{67}H_{93}N_{11}O_{14}S$

RPHPLC: Jupiter C4 300 Å;
gradient 1: t_{ret} = 18.7 min

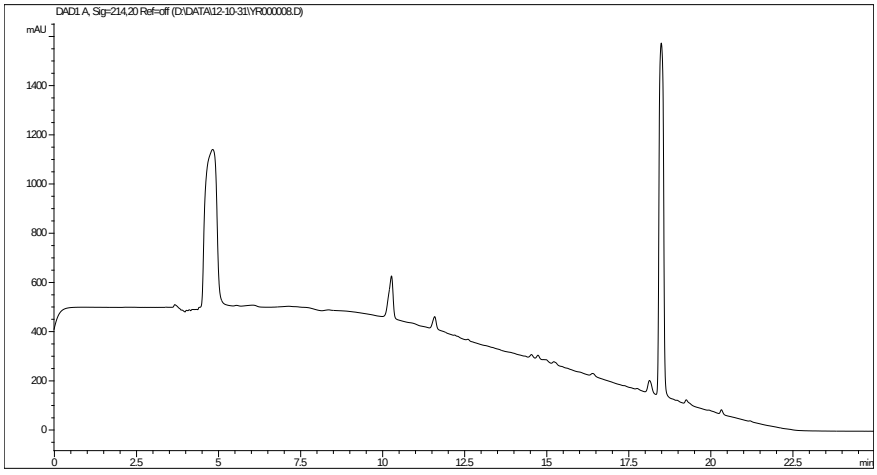


Figure 11.6.1: RP-HPLC analysis using Jupiter C4 300 Å; gradient: 0 to 100 % increase of B (versus A) in 15 min: t_{ret} = 18.7 min.

Maldi-TOF-MS (DHB): Exact identity of C-terminal functionality remains uncertain due to presumable Maldi-induced hydrolysis of labile species. Complementing data of above chaperone cleavage experiments: calcd. EM 1307.7; MW 1308.6; found *m/z* 1308.5 $[M + H]^+$, 1330.5 $[M + Na]^+$, 1346.5 $[M + K]^+$, 1208.5 $[M - Boc + H]^+$, 1056.5 $[M - Pbf + H]^+$

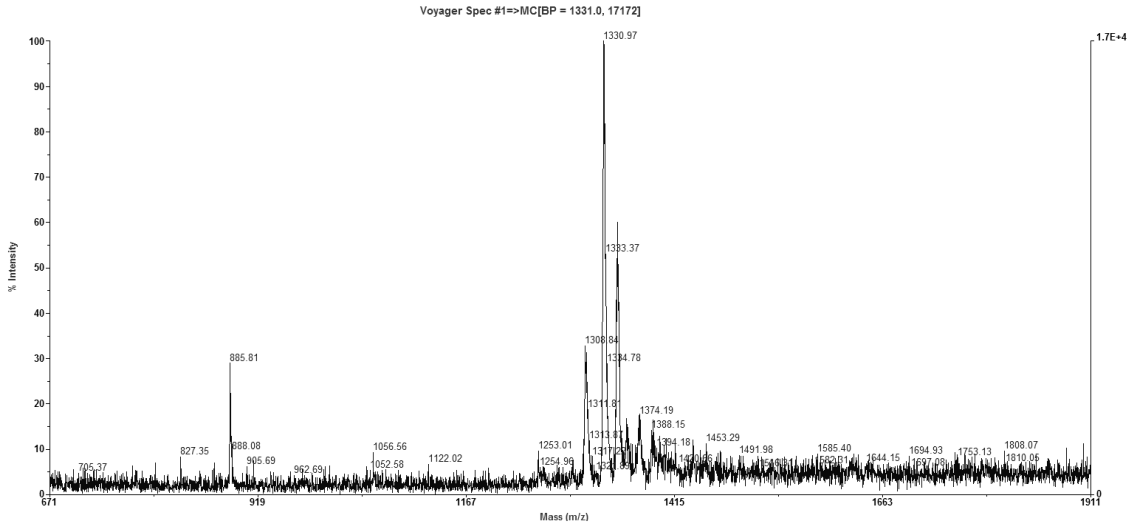
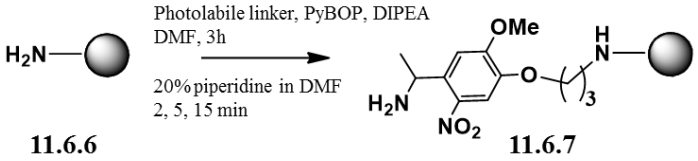


Figure 11.6.2: MALDI-TOF Spectrum of crude compound **11.6.5**. Calculated m/z (100%) = 1307.7; m/z (74.1 %) = 1308.67. Found: 1308.5 = M + H+, 1330.5 [M + Na]$^+$, 1346.5 [M + K]$^+$, 1208.5 [M - Boc + H]$^+$, 1056.5 [M - Pbf + H]$^+$ MALDI-TOF Spectrum of crude compound 5. Calculated m/z (100%) = 1307.7; m/z (74.1 %) = 1308.67. Found: 1308.8 = M + H+, 1330.9 [M + Na]$^+$.

11.6.3 RRTR to C12 to obtain intermediate peptidosteroid conjugate

Optimized RRTR prototype validation for decoration of the Tentagel-bound, steroid scaffold

Synthesis of photolinker-bearing TentaGel resin

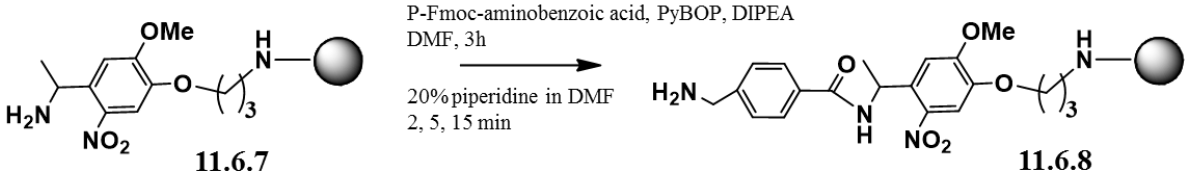


11.6.6 **11.6.7**

In a large glass reactor, TentaGel resin **11.6.6** (0.14 mmol based on manufacturer's loading of TentaGel precursor, synthesized as detailed above) was immersed in dry DMF (3 mL) and NHFmoc-protected Holme's Photolabile linker (218.6 mg, 0.42 mmol, 3 eq.), dry DIPEA (146 µL, 0.84 mmol, 6 eq.) and PyBOP (218.5, 0.42 mmol, 3 eq.) were consecutively added. Solid reagents were readily dissolved, generating an intense orange mixture. Additional dry DMF (2 mL) completed the reaction mixture, which was flushed with argon and shielded from light. Coupling reaction proceeded for 3 hours at room temperature, excess reagents and solvent were removed by filtration under reduced pressure and resin was washed with DMF, ACN and DCM.

Synthesis of C24 *p*-aminomethylbenzyl amide tagged acceptor construct on Tentagel resin

Coupling of *p*-(NHFmoc)aminomethylbenzoic acid to photocleavable construct and NHFmoc peprotection



In a large glass reactor, pale orange-yellow, photolinker-bearing TentaGel resin **11.6.7** (0.14 mmol) was immersed in dry DMF (3 mL) and NHFmoc-protected *p*-aminomethylbenzoic acid (156.8 mg, 0.42 mmol, 3 eq), dry DIPEA (140 µL, 0.84 mmol, 6 eq) and PyBOP (218.7, 0.42 mmol, 3 eq) were consecutively added. Solid reagents were readily dissolved, generating an intense orange mixture. Additional dry DMF (2 mL) completed the reaction mixture, which was flushed with argon and shielded from light. Coupling reaction proceeded for 3 hours at room temperature, excess reagents and solvent were removed by filtration under reduced pressure and resin was washed with DMF, ACN and DCM.

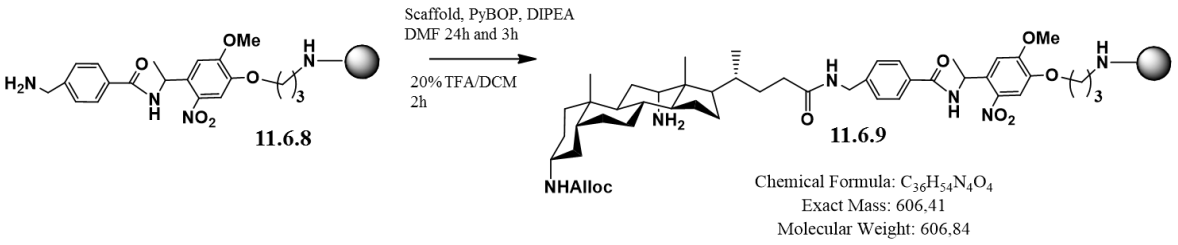
Coupling was repeated (2 ½ hour) after washing the resin with additional dry DMF.

TNBS test: colourless (pale yellow resin) for prolonged time (> 1 hour)

Removal of the NHFmoc-protectant proceeded by the usual, triple treatment with piperidine/ DMF (20 % v/v, 5 mL) as described earlier.

TNBS test: intense orange-red in 5 minutes

Coupling of steroid scaffold and C12-NHBoc deprotection ➔ 11.6.9



Adopting the procedure described above, steroid scaffold was attached to the resulting construct **11.6.8** by double coupling at room temperature, once overnight and once for 3 hours.

TNBS test: colourless

Subjecting resulting resin **11.6.9** to TFA/DCM (20 % v/v, 5 mL) for 2 hours, C12-NHBoc group was removed, generating the desired TentaGel-bound acceptor construct. Final, orange-brownish resin was washed with DCM, ACN and DCM and thoroughly dried under high vacuum.

H_2N-C12, AllocHN-C3-steroid-C24-'ABA'-$CONH_2$ $C_{36}H_{54}N_4O_4$

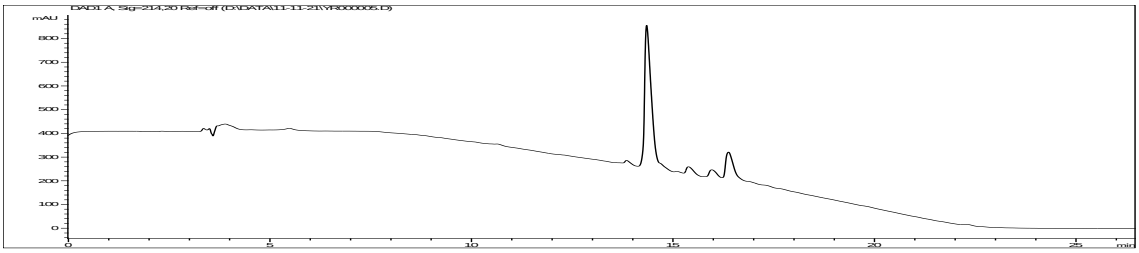


Figure 11.6.3: RP-HPLC analysis using Jupiter C4 300 Å; gradient: 0 to 100 % increase of B (versus A) in 15 min: t_{ret} = 14.5 min.

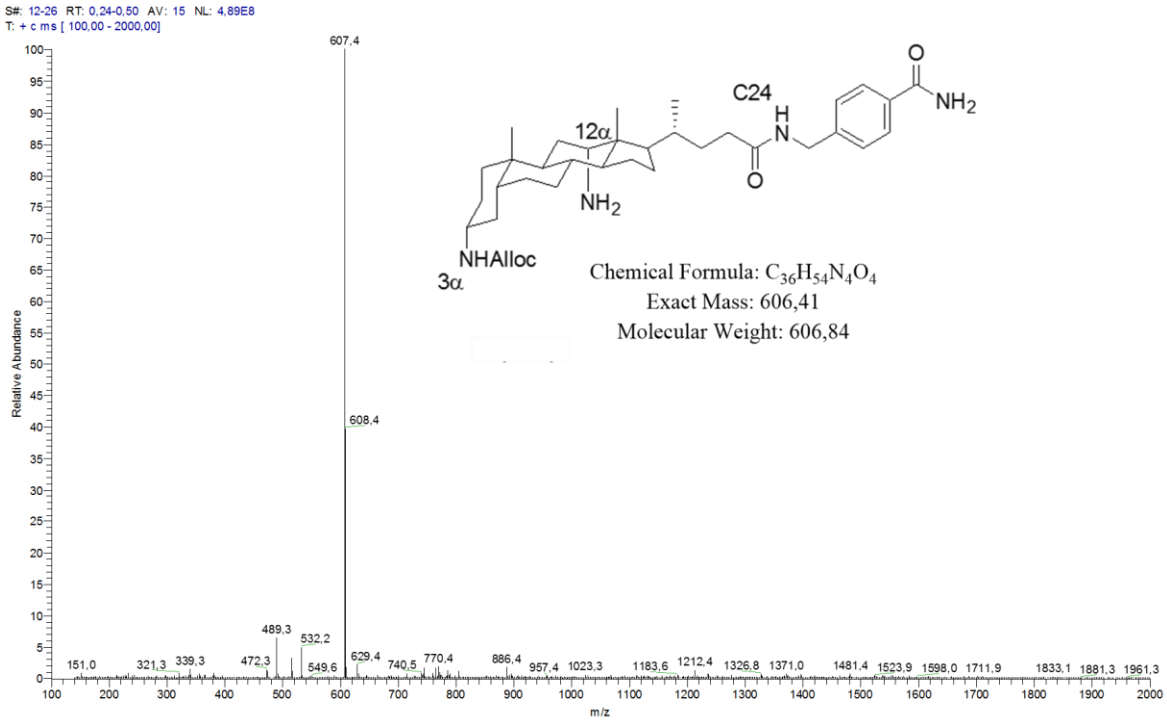
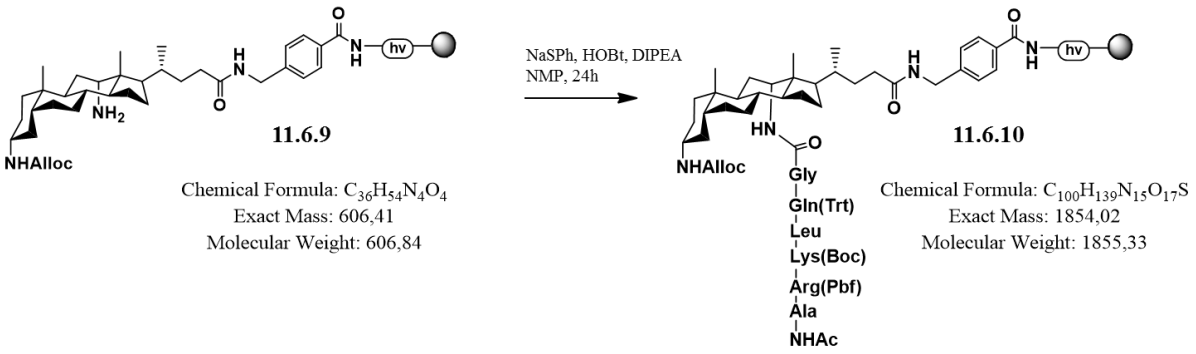


Figure 11.6.4: ESI-MS of RPHPLC-isolated (Luna C18 100 Å; gradient 1) main peak: calcd. EM 606.4, MW 606.8; found *m/z* (% rel. int.) 607.4 (100) $[M + H]^+$, 629.4 (2) $[M + Na]^+$.

HOBt-mediated transfer of side-chain protected AcHN-ARKLQG-Ellman (4) to $\underline{H_2N}$-C12-, AllocHN-C3-steroid-C24-'ABA'-Tentagel.



Bearing the photocleavable, UV-active steroid scaffold acceptor, TentaGel resin **11.6.9** (20.0 mg, theoretical loading calcd. from TentaGel: 0.23 mmol/g, 0.0046 mmol) was weighed in a SYRO tube (2 mL). Providing the shuttled peptide with preloaded, C-terminal glycine residue, activated safety-catch donor resin **11.6.4** (124.3 mg, 0.184 mmol/g, 0.023 mmol, 5 eq) and fresh, dry NMP (700 µL) were added. After gently swirling the resin mixture, NaSPh (30.4 mg, dried, 0.23 mmol, 50 eq), HOBt (31.1 mg, 0.23 mmol, 50 eq) and dry DIPEA (38 µL, 0.23 mmol, 50 eq) were added, the reactor flushed with argon and the reaction vortexed overnight (24 hours) at room temperature. Excess reagents and solvent were removed by filtration under reduced pressure and the filtrate was isolated in a round-bound flask under argon atmosphere. A sample was diluted with ACN (1/12) and analyzed.

Discarding the washing batches, beads were further rinsed with NMP, ACN and DCM. The resin mixture (containing ~86 wt% inert safety-catch beads) was sampled in a glass tube and desired peptidosteroid conjugate **(10)** photolytically released in ACN. Filtrates were diluted in ACN ($1/12^{th}$) and together with cleaved samples analyzed by RPHPLC and Maldi-TOF-MS.

Side-chain protected AcHNARKLQG-C12, AllocHN-C3-steroid-C24-'ABA'-$CONH_2$ $C_{103}H_{145}N_{15}O_{17}S$

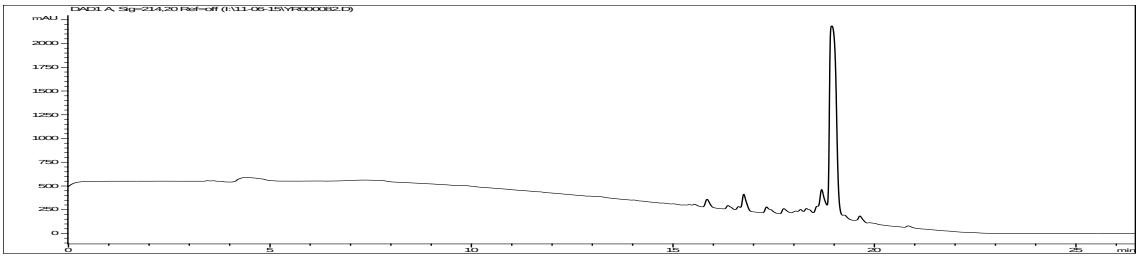


Figure 11.6.5: RPHPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 19 min.

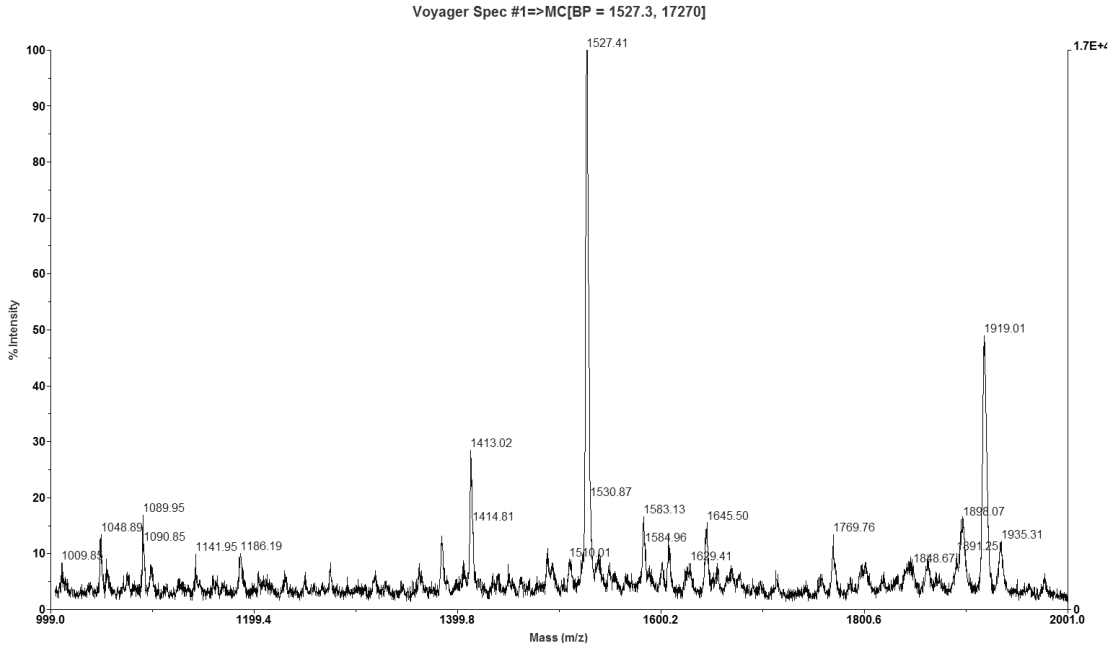
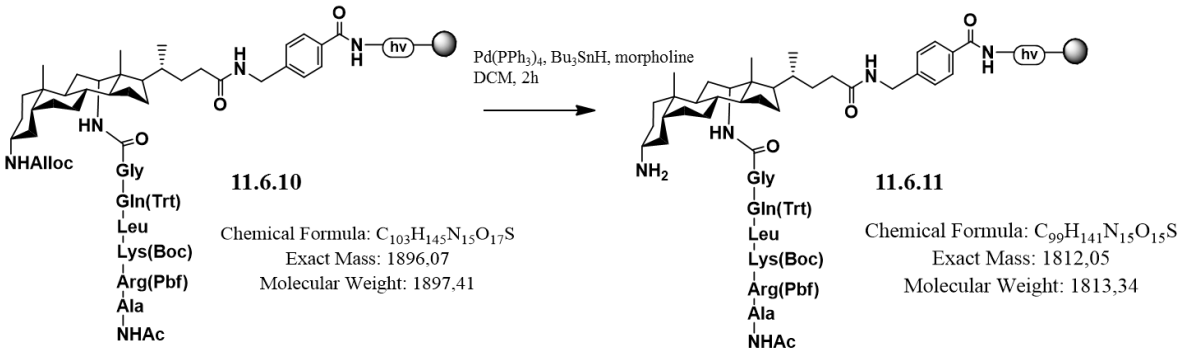


Figure 11.6.6: MALDI of crude compound: calcd. EM 1896.07, MW 1897.41; Found m/z (% rel. int.) 1919 $[M + Na]^+$, 1527 $[M\text{-}Pbf\text{-}Boc\text{-}18 + H]^+$.

11.6.4 RRTR to C3 and final deprotection

Alloc deprotection



Prior to alloc deprotection, resin **11.6.10** (40 mg, 0.23 mmol/g, 0.092 mmol) was swollen with DCM. After the swelling of the resin, dry DCM was added to the reactor, followed by the addition of $Pd(PPh_3)_4$ (106.3 mg, 1 eq.), Bu_3SnH (247 µL, 10 eq.) and morpholine (716 µL, 90 eq.). Yellow/orange mixtures were flushed with argon, shielded from light and agitated for 2 hours.

Excess reagents and solvent were removed under reduced pressure and the resin was washed with DCM, ACN and DCM.

Side-chain protected AcHNARKLQG-C12, H_2N-C3-steroid-C24-'ABA'-$CONH_2$ $C_{99}H_{141}N_{15}O_{15}S$

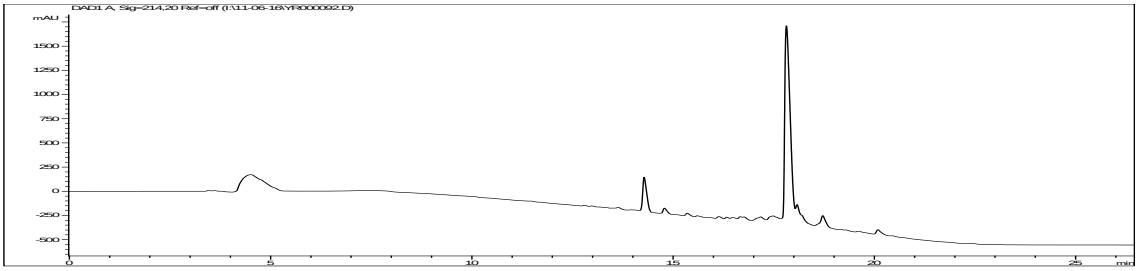


Figure 11.6.7: RPHPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 18 min.

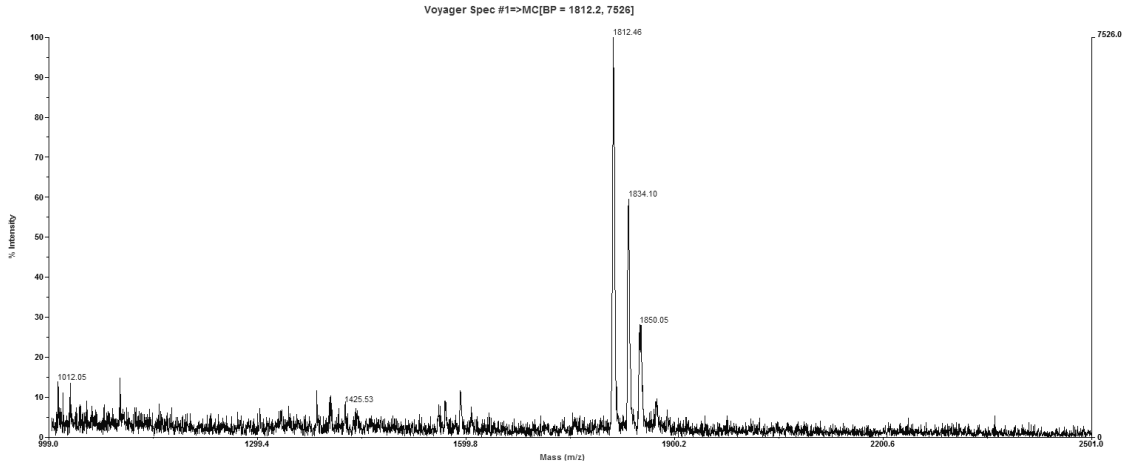
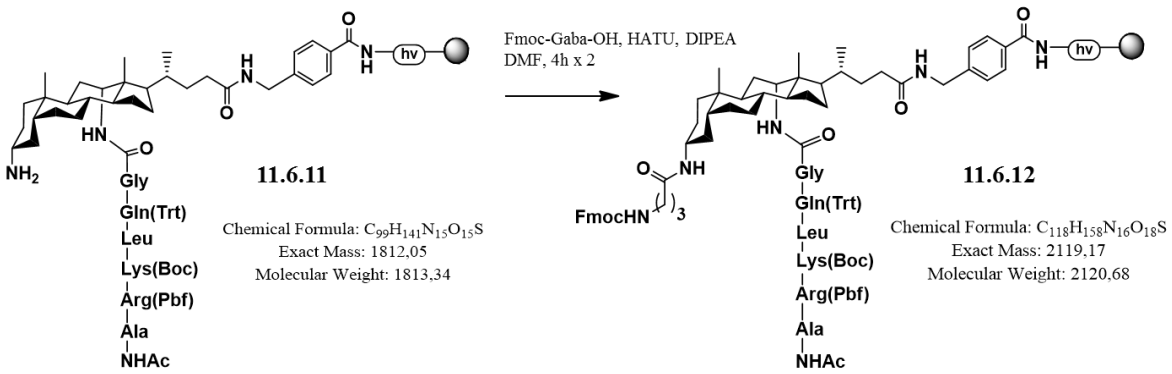


Figure 11.6.8: MALDI of crude compound: calcd. EM 1811.05, MW 1812.34; Found m/z (% rel. int.) 1812 $[M + H]^+$, 1834 $[M + Na]^+$, 1850 $[M + K]^+$.

Coupling of Fmoc-Gaba-OH at C3



In a glass reactor, scaffold-bearing resin **11.6.11** (30 mg, loading 0.27673 mmol/g) was swollen in dry DMF. After the swell process was repeated three times for 2 minutes, it is suspended in dry DMF (1 mL) and dry DIPEA (5 µL, 0,0276 mmol, 4 eq), Fmoc-Gaba-OH (8,98 mg, 0,0276 mmol, 4 eq.) and HATU (10,49 mg, 0,0276 mmol, 4 eq.) are successively added and shaked for homogeneity. The mixture is flushed with argon, shielded from light and gently shaken for four hours at room temperature.

Excess of reagents were removed by filtration. The resin was washed with DMF, ACN, MeOH and DCM.

The coupling is repeated twice to get the reaction into completion.

Side-chain protected AcHNARKLQG-C12, Fmoc-HN-GABA-HN-C3-steroid-C24-'ABA'-$CONH_2$ $C_{118}H_{158}N_{16}O_{18}S$.

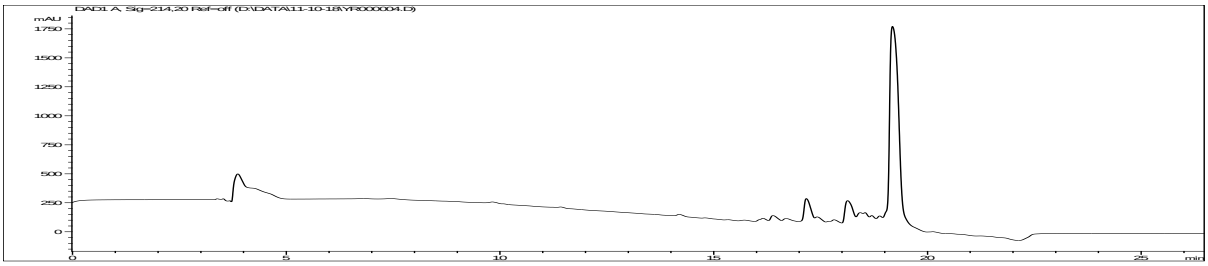


Figure 11.6.9: RP-HPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 19.5 min.

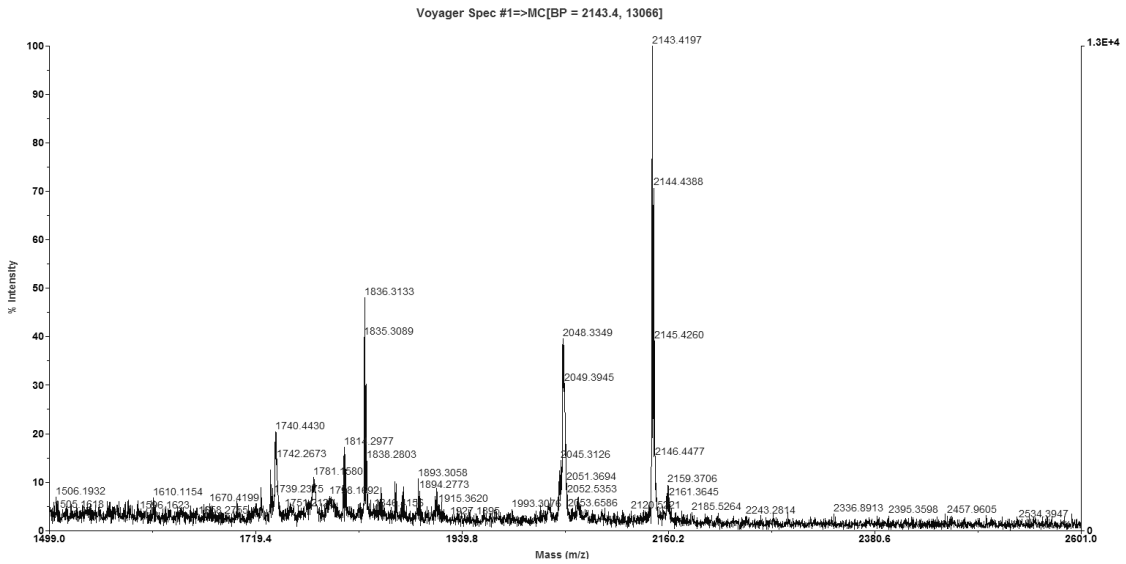
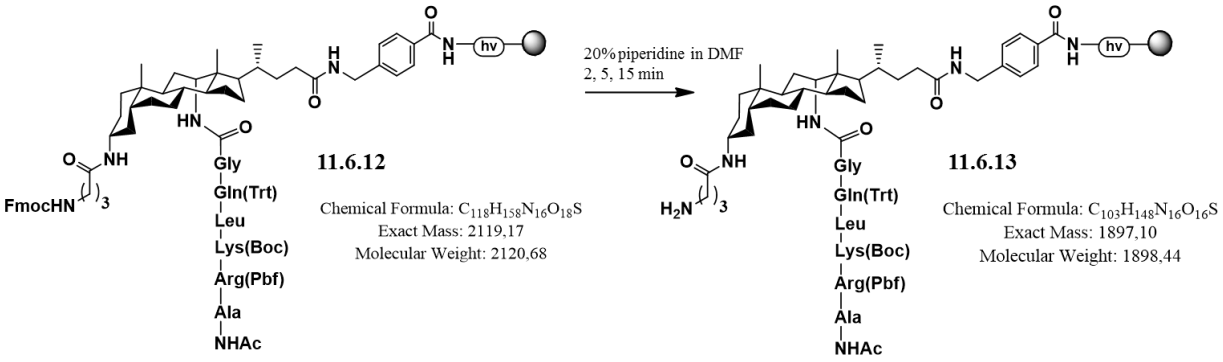


Figure 11.6.10: MALDI of crude compound: calcd. EM 2119.17, MW 2120.68; Found m/z (% rel. int.) 2143.42 $[M + Na]^+$.

Fmoc deprotection



The amine from the Gaba linker was deprotected with a mixture of 20% piperidine in DMF (v/v). 1 mL was added for 30 mg of resin. The procedure is repeated 3 times for 2, 5 and 15 minutes. After each repetition, the solution is filtered and the resin is washed throroughly with DMF. The final washing step is achieved with DMF, ACN and DCM.

Side-chain protected AcHNARKLQG-C12, H_2N-GABA-HN-C3-steroid-C24-'ABA'-$CONH_2$ $C_{103}H_{148}N_{16}O_{16}S$.

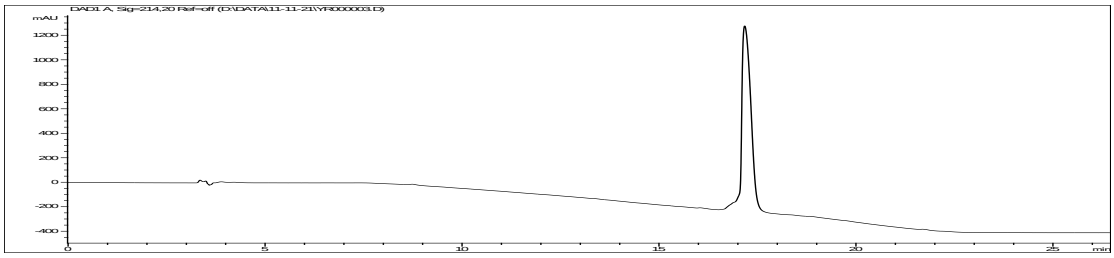


Figure 11.6.11: RPHPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 17 min.

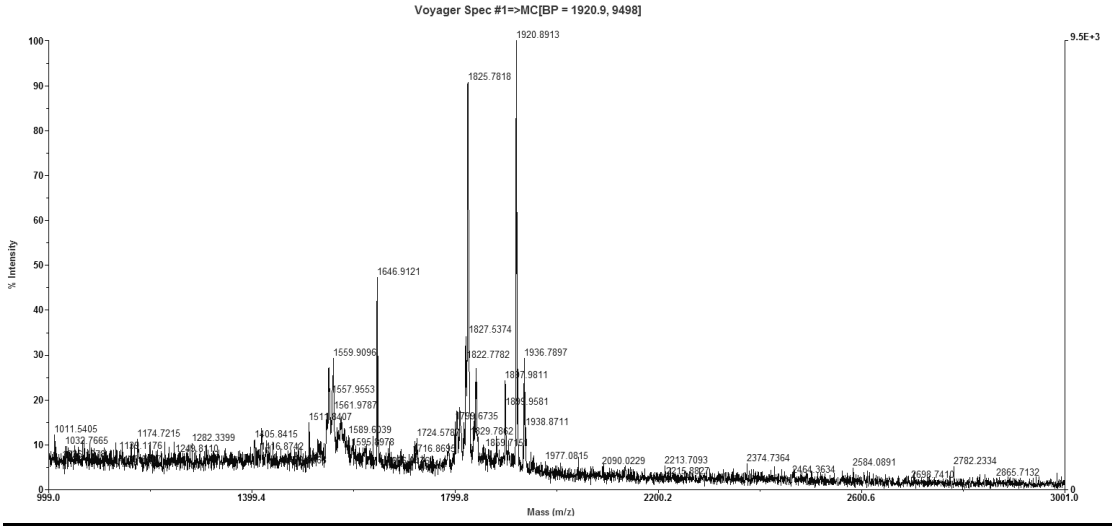
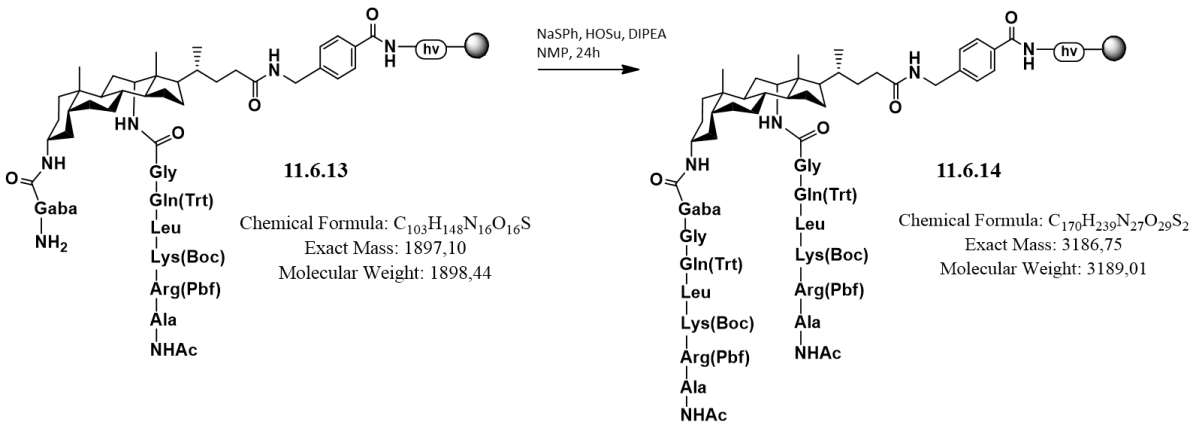


Figure 11.6.12: MALDI of crude compound: calcd. EM 1897.10, MW 1898.44; Found m/z (% rel. int.) 1920.89 [M + Na]$^+$.

RRTR to C3



TentaGel acceptor resin **11.6.13** (5 mg, 0.00115 mmol, 0.23 mmol/g) was transferred to a filter reactor (2 mL). Activated donor-bearing safety-catch resin **11.6.4** (32.9 mg, 0.00575 mmol, 5 eq.) and dry DMF was added, followed by NaSPh (7.6 mg, 0.0575 mmol, 50 eq.), dry DIPEA (20 µL, 0,115 mmol, 100 eq) and HOSu (6.62 mg, 0.0575 mmol, 50 eq.). The resulting mixture was flushed with argon and vortexed overnight (24 hours) at room temperature, shielded from light. Resin was purified by filtration under reduced pressure and washed with DMF, ACN and DCM.

Table 11.6.1: Experiments without Gaba linker at C3

Chaperone	Equivalents	DIPEA	Donor resin
HOBt	50	50	5
HOAt	50	50	5
HOSu	50	50	5
Oxyma pure	50	50	5

Table 11.6.2: Experiments with Gaba at C3

Chaperone	Equivalents	DIPEA	Donor resin
HOBt	50	100	5
HOAt	50	100	5
HOSu	50	100	5
Oxyma pure	50	100	5

Side-chain protected AcHNARKLQG-C12, AcHNARKLQG-HN-GABA-HN-C3-steroid-C24-'ABA'-$CONH_2$ $C_{170}H_{239}N_{27}O_{29}S_2$.

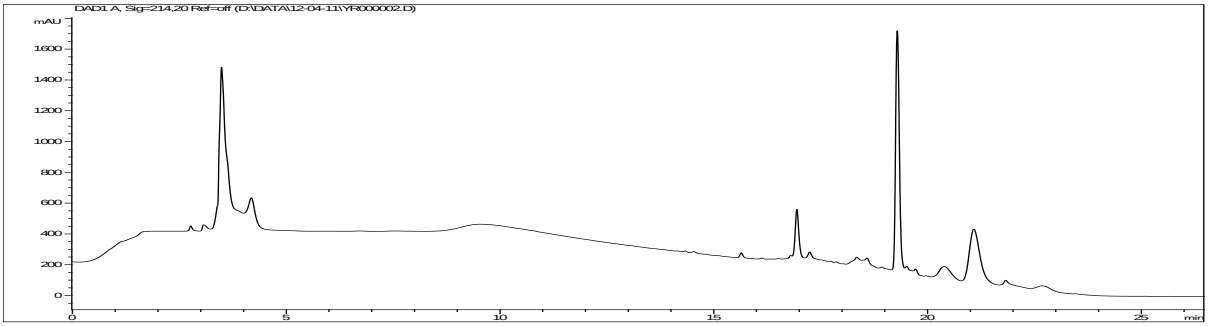


Figure 11.6.13: RPHPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 19.5 min.

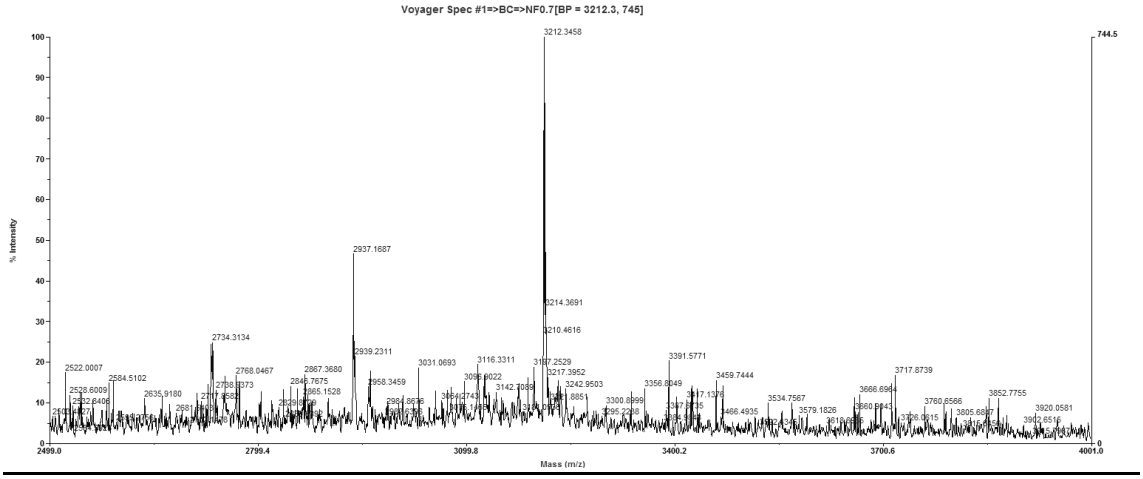
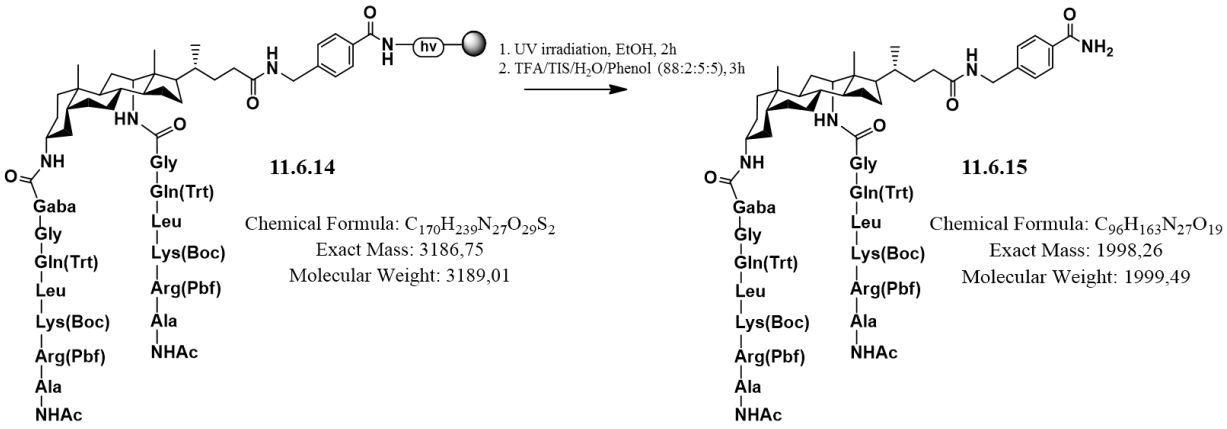


Figure 11.6.14: MALDI of crude compound: calcd. EM 3186.75, MW 3189.01; Found m/z (% rel. int.) 3212.34 [M + Na]$^+$.

Final cleavage and deprotection



Resin **11.6.14** in EtOH is irradiated under UV light for 2 hours at room temperature. Then, the supernatant solution is transferred to a flask and the solvent is evaporated. Side chain protecting groups are removed by adding to the cleaved compound a cocktail with TFA:TIS:H_2O (95:2.5:2.5) and shaking for 3 hours. The solvent is removed under reduced pressure and the obtained residue purified by centrifugation in cold MTBE to get the desired final compound **11.6.15**.

Side-chain deprotected AcHNARKLQG-C12, AcHNARKLQG-HN-GABA-HN-C3-steroid-C24-'ABA'-$CONH_2$ **11.6.15** $C_{96}H_{163}N_{27}O_{29}$.

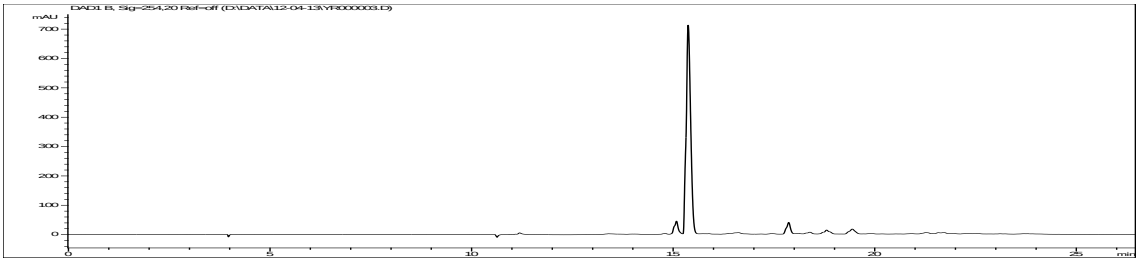


Figure 11.6.15: RPHPLC analysis using Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min: t_{ret} = 15.5 min.

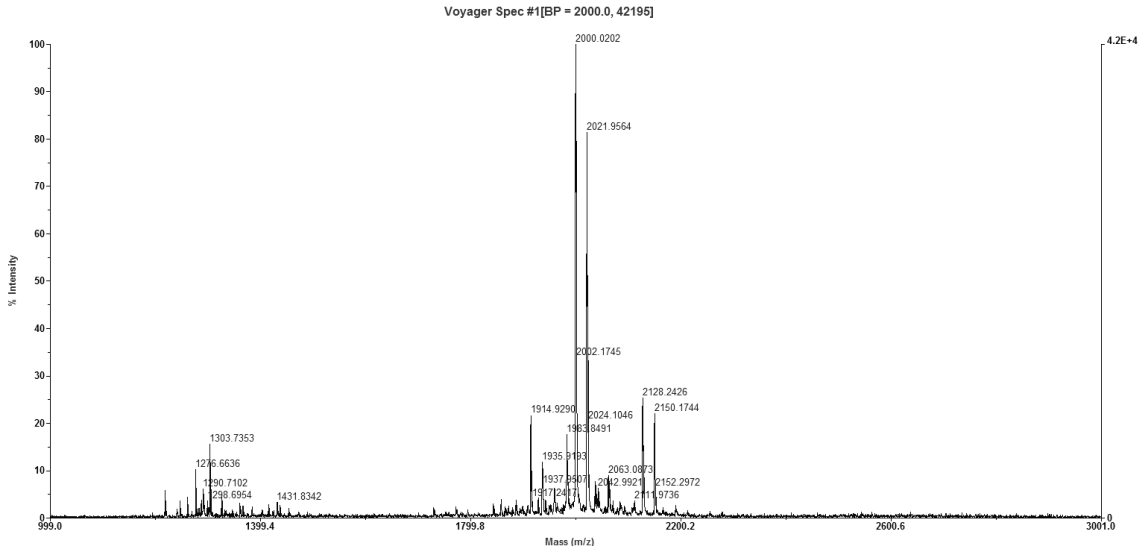
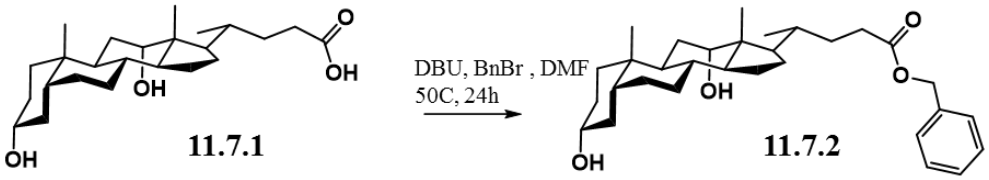


Figure 11.6.16: MALDI of crude compound: calcd. EM 1998.26, MW 1999.49; Found m/z (% rel. int.) 2000.02 $[M + H]^+$.

CHAPTER 11.7

11.7.1 Synthesis of scaffold for DNA cleavage

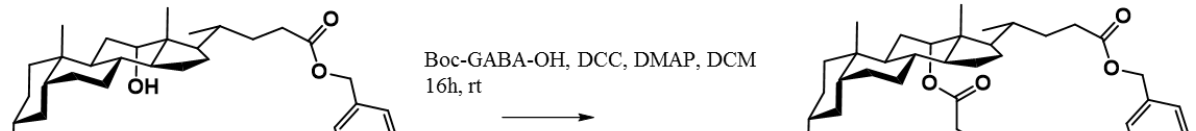
Benzyl 3α, 12α-dihydroxy-5β-cholan-24-oate (11.7.2). For synthesis, refer to **11.2.1**.



Chemical Formula: $C_{24}H_{40}O_4$
Exact Mass: 392,29
Molecular Weight: 392,57

Chemical Formula: $C_{31}H_{46}O_4$
Exact Mass: 482,34
Molecular Weight: 482,69

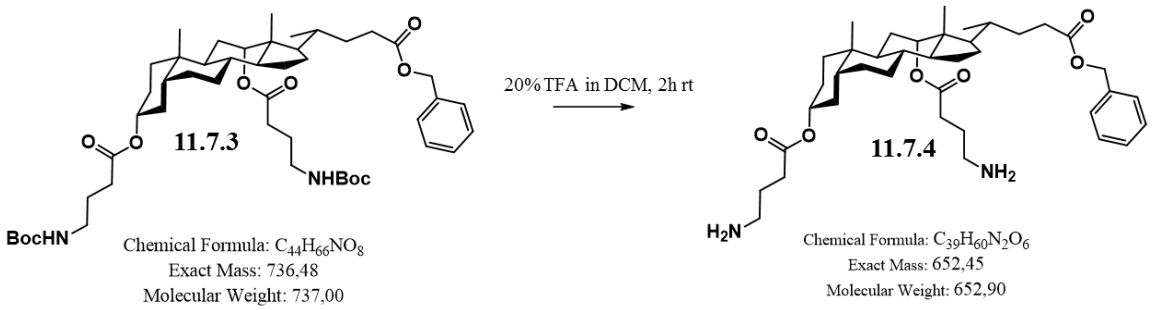
Synthesis of intermediate 11.7.3. Compound **11.7.2** (300 mg, 0.622 mmol) was dissolved in 2 mL dry DCM. Then a solution of Boc-GABA-OH (379.5 mg, 1.86 mmol), DCC (385.2 mg, 1.86 mmol) and DMAP (304.1 mg, 2.49 mmol) in dry DCM (3 mL) was added. The reaction mixture was stirred for 16h at room temperature. The precipitate formed was filtered off and the solvent was evaporated under vacuo. The residue was purified by column chromatography (silica gel DCM/MeOH 96:4; Rf: 0.40) to give compound **11.7.3** as a yellow oil (421.3 mg, 90% yield). ^{1}H-NMR (300 MHz, CD_2Cl_2); δ 7.38 − 7.36 (m, 5 H), 5.15 (m, 3H), 4.65 (br, m, 1 H), 3.15 (m, 4H), 1.45 (s, 18H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H).



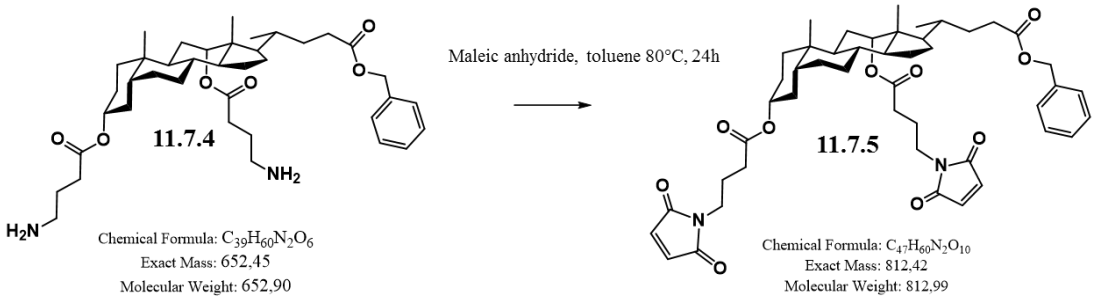
Chemical Formula: $C_{31}H_{46}O_4$
Exact Mass: 482,34
Molecular Weight: 482,69

Chemical Formula: $C_{44}H_{66}NO_8$
Exact Mass: 736,48
Molecular Weight: 737,00

Synthesis of intermediate 11.7.4. Compound **11.7.3** (300 mg, 0.407 mmol) was dissolved in 5 mL DCM. 1mL TFA was added to the solution and stirred for 2h. After completion, the solvent was evaporated in vacuo and a transparent oil was obtained (262.9 mg, 99% yield). ^{1}H-NMR (300 MHz, CD_2Cl_2); δ 7.38 − 7.36 (m, 5 H), 5.15 (m, 3H), 4.65 (br, m, 1 H), 3.45 (m, 4H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H).



Synthesis of intermediate 11.7.5. Compound **11.7.4** (250 mg, 0.383 mmol) was dissolved in 6 mL dry toluene. Maleic anhydride (268.52 mg, 1.532 mmol) was added to the solution and the mixture stirred at 80°C for 24h. The solvent was evaporated and the crude compound purified by column chromatography (fine silica gel EtOAc/Hexane 1:4; Rf: 0.60) to give compound **11.7.5** as a transparent oil (217.8 mg, 70% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 6.70 (m, 4H), 5.12 (d, 1H, J = 13.56 Hz), 5.07 (d, 1H, J = 13.75 Hz), 4.65 (br, m, 1 H), 3.55 (m, 4H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ^{13}C NMR (125 MHz, $CDCl_3$): δ = 173.71 (COOR-C24), 171.82 (4C maleimides), 171.57 (COOR-C3), 170.45 (COOR-C12), 135.80 (C - benzyl group), 133.86 (2CH maleimides), 133.83 (2CH maleimides), 128.27 (2CH-benzyl ring), 127.99 (2CH- benzyl ring), 127.92 (CH-benzyl ring), 75.84 (CH ester), 74.09 (CH ester), 65.85 (CH_2-benzyl ester), 49.03 (CH), 47.26 (CH), 44.80 (C), 41.54 (CH), 36.85 (C), 35.40 (CH), 34.20 (CH), 34.15 (CH_2), 34.11 (CH), 33.78 (CH_2), 31.60 (CH_2), 31.53 (CH_2), 30.93 (CH_2), 30.60 (CH_2), 26.90 (CH_2), 26.75 (CH_2), 26.65 (CH_2), 26.43 (CH_2), 26.15 (CH_2), 25.32 (CH_2), 23.86 (CH_2), 23.60 (CH_2), 23.18 (CH_2), 22.76 (19-CH_3), 17.22 (21 – CH_3), 14.71 (CH_2), 12.11 (18-CH_3). ES-MS m/z (% rel. int.) calcd. 812.4 (100), found 830.5 (100) [M+NH_4^+]. HR-MS (ES) m/z calcd. for $C_{35}H_{48}N_6O_6$ + NH_4^+ 812.4248, found 830.4607 (M + NH_4^+; Δ = 2.6 ppm).



11.7.2 NMR Spectra for steroid template molecules:

Compound 11.7.3:

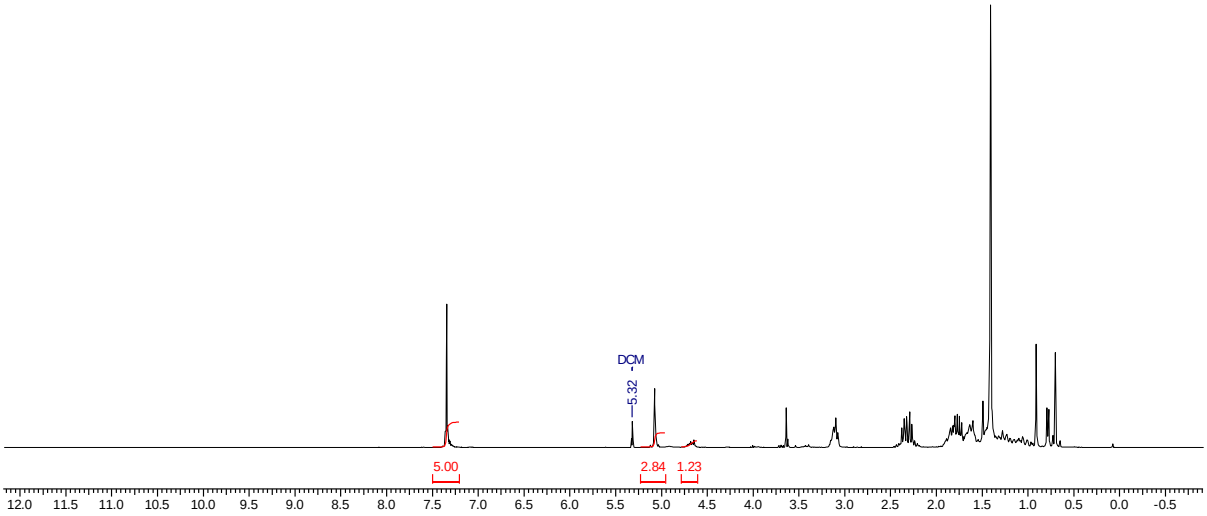


Figure 11.7.1. ^{1}H NMR spectrum of compound **11.7.3**.

Compound 11.7.4:

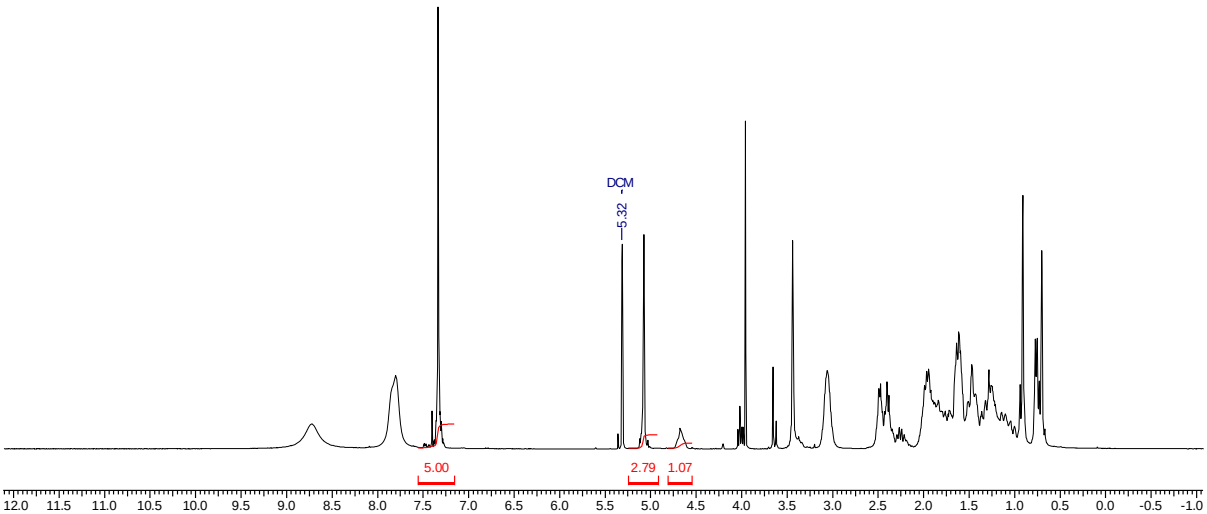


Figure 11.7.2. ^{1}H NMR spectrum of compound **11.7.4**.

Compound 11.7.5:

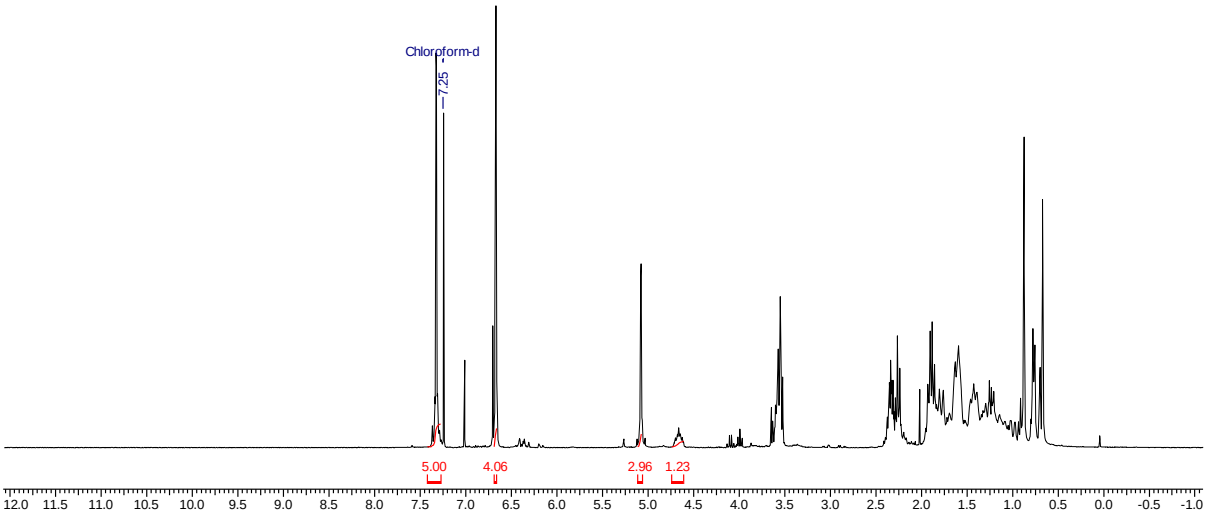


Figure 11.7.3. ^{1}H NMR spectrum of compound **11.7.5**.

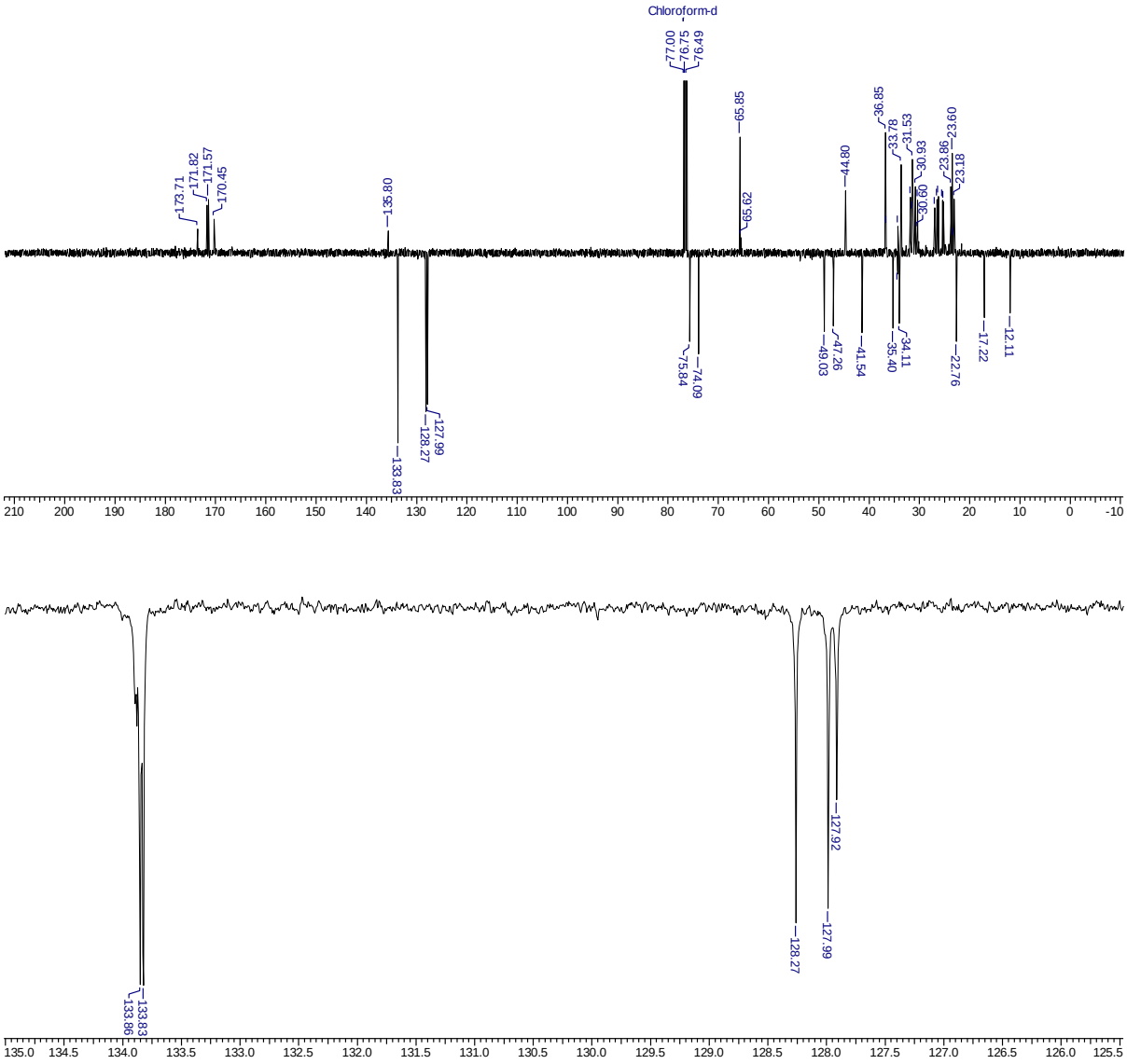
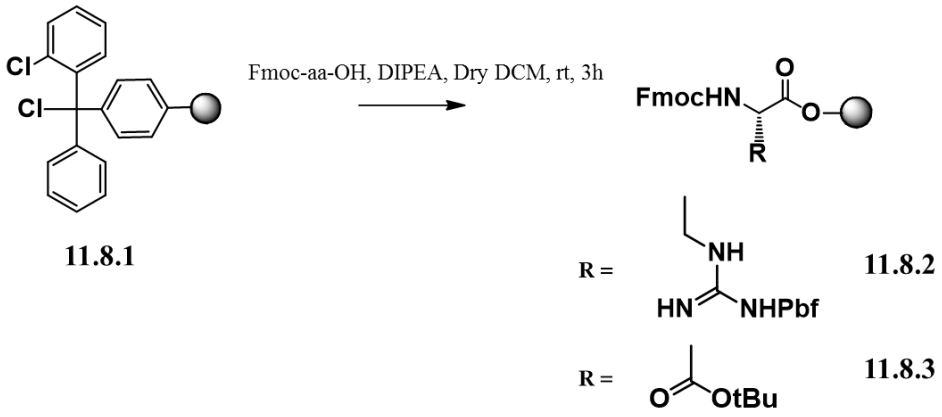


Figure 11.7.4. ^{13}C NMR spectrum of compound **11.7.5**.

CHAPTER 11.8

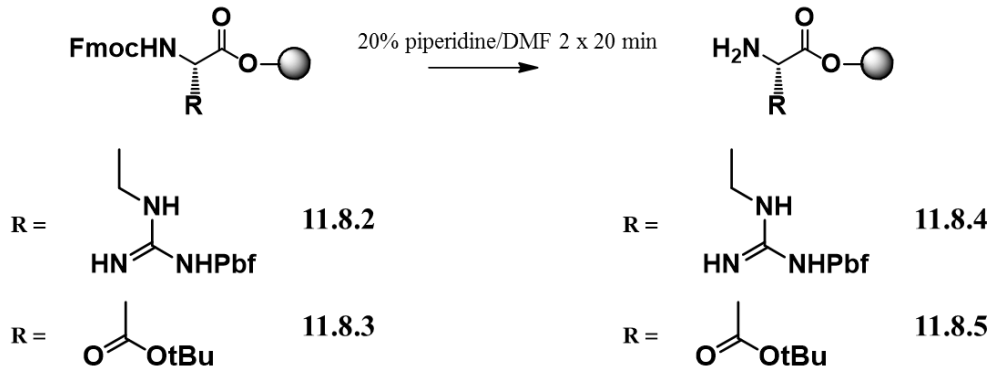
11.8.1 Synthesis of synthetic receptor for EDCs

Coupling of first amino acid



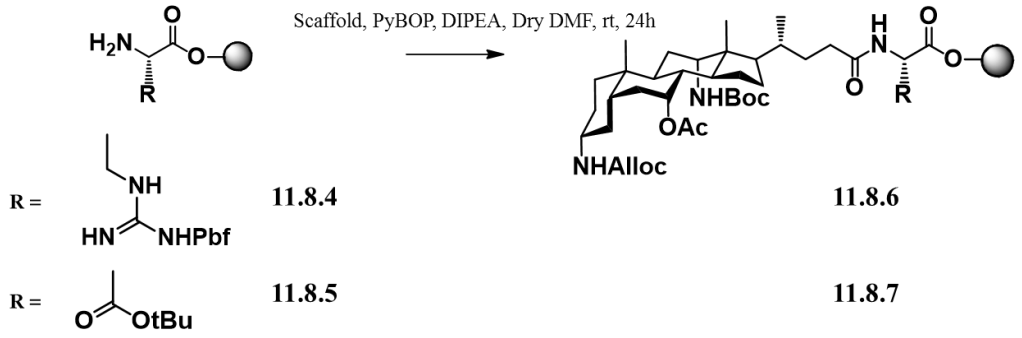
2-chlorotrityl chloride resin **11.8.1** (0.23 g, 1.55 mmol/g) were swollen in dichloromethane (dry, 10 mL/g) prior to the loading of the first amino acid. Fmoc-Glu(tBu)-OH (0.30 g, 0.71 mmol) or Fmoc-Arg(Pbf)-OH (0.46 g, 0.71 mmol) were dissolved in dry DCM with N,N'-diisopropylethylamine (DIPEA) (0.74 mL, 4.28 mmol) and the reactions were shaken for 3 h. The resins were then washed with 3 x DCM, 3 x DMF, 3 x DCM and dried under vacuum for 2 h. The loading was calculated to be 0.55 mmol/g, consistent with a coupling yield of 71%. The resins were then capped with methanol/DIPEA/DCM = 2:1:17 by shaking for 2 h.

Fmoc deprotection for the first amino acid



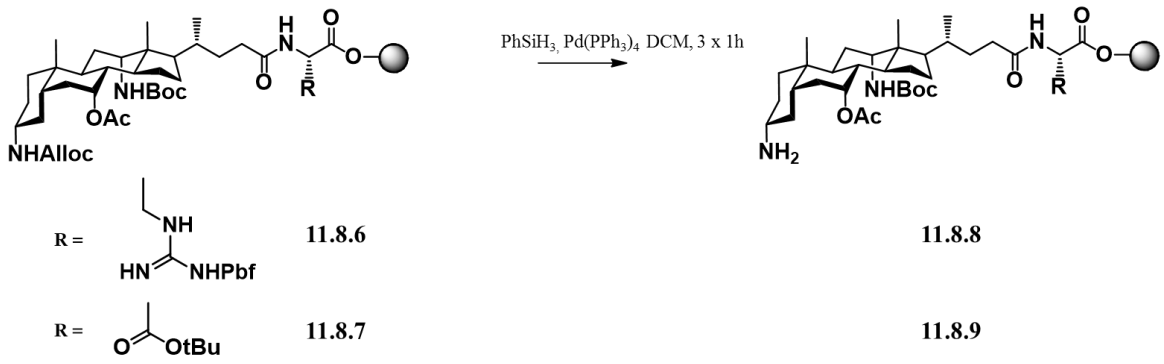
The resins **11.8.2** and **11.8.3** were swollen in DMF prior to Fmoc deprotection. A solution of 20% piperidine in DMF was added to the resins and the mixtures were shaken for 2 x 30 min. The solutions were then washed with 3 x DMF, 3 x methanol, 3 x DMF.

Coupling of the scaffold



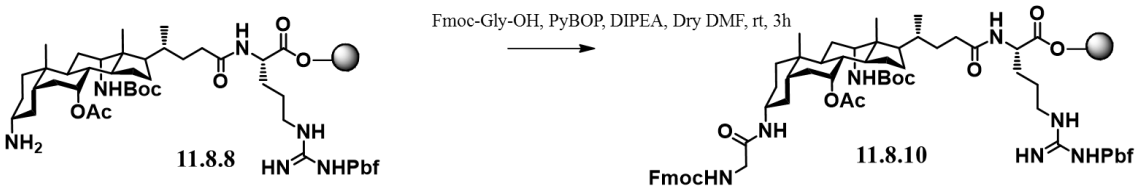
At a suspension of resins **11.8.4** and **11.8.5** in DMF were added the dipodal scaffold (0.23 g, 0.37 mmol, 0.5 M), PyBOP (0.19 mg, 0.37 mmol, 0.5 M) and DIEA (0.13 mL, 0.74 mmol, 2 M). The mixtures were shaken at room temperature for 24 h, after which the resins were washed with 3 x DMF, 3 x methanol, 3 x DMF.

Alloc deprotection

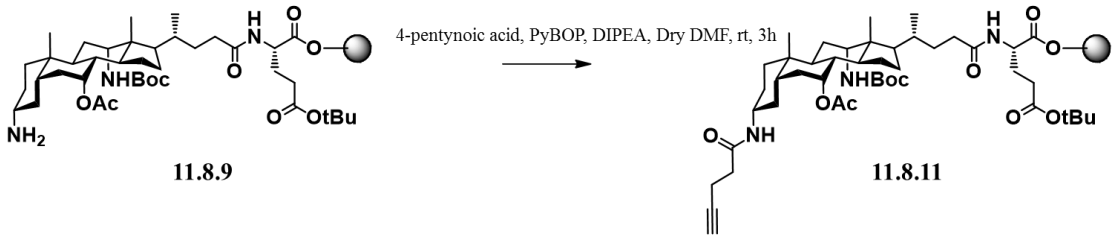


Resins **11.8.6** and **11.8.7** (0.22 g, 0.55 mmol/g) were swollen and thoroughly washed with DCM under a stream of argon. The resins were suspended in DCM (6.7 mL) then were loaded the scavenger $PhSiH_3$ (0.46 mL, 3.7 mmol) and the catalyst $Pd(PPh_3)_4$ (0.01 g, 0.01 mmol). Yellow/orange mixtures were flushed with argon, shielded from light and shaked at room temperature for 1 h under anhydrous conditions, after which the resins were washed with DCM and the reactions were repeated twice. At the end of the Alloc removal the resins were washed with 3 x DMF, 3 x methanol, 3 x DMF.

Coupling of Fmoc-Gly-OH and 4-pentynoic acid



To a suspension of the Alloc deprotected resin **11.8.8** (0.22 g, 0.55 mmol/g) in DMF were added Fmoc-Gly-OH (0.18 g, 0.62 mmol, 0.5 M), PyBOP (0.32 g, 0.62 mmol, 0.5 M) and DIEA (0.21 mL, 1.2 mmol, 2 M). The mixture was shaken at room temperature for 3 h.



To resin **11.8.9** (0.095 g, 0.96 mmol/g) in DMF were added 4-pentynoic acid (0.035 g, 0.353 mmol, 0.5 M), PyBOP (0.184 g, 0.353 mmol, 0.5 M) and DIEA (0.123 mL, 0.71 mmol, 2 M). The mixture was shaken at room temperature for 2 h. A small amount of resin **11.8.11** was cleaved in mild condition for LC-MS analysis: t_R 14.26 min; exact mass calculated: 813.51; found m/z 814.40 $[M + H]^+$ and 714.40 $[M - Boc]^+$.

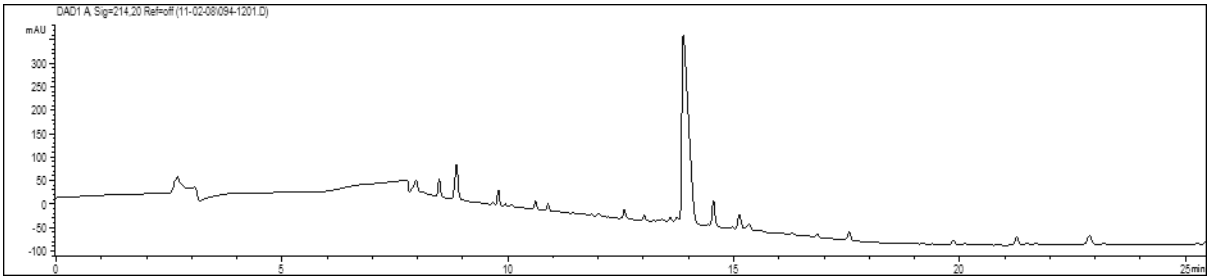


Figure 11.8.1: HPLC trace from LC-MS of compound **11.8.11** (0-100% ACN in 15 min on Luna C18 250 x 4.6 mm, 5 µm at 35°C).

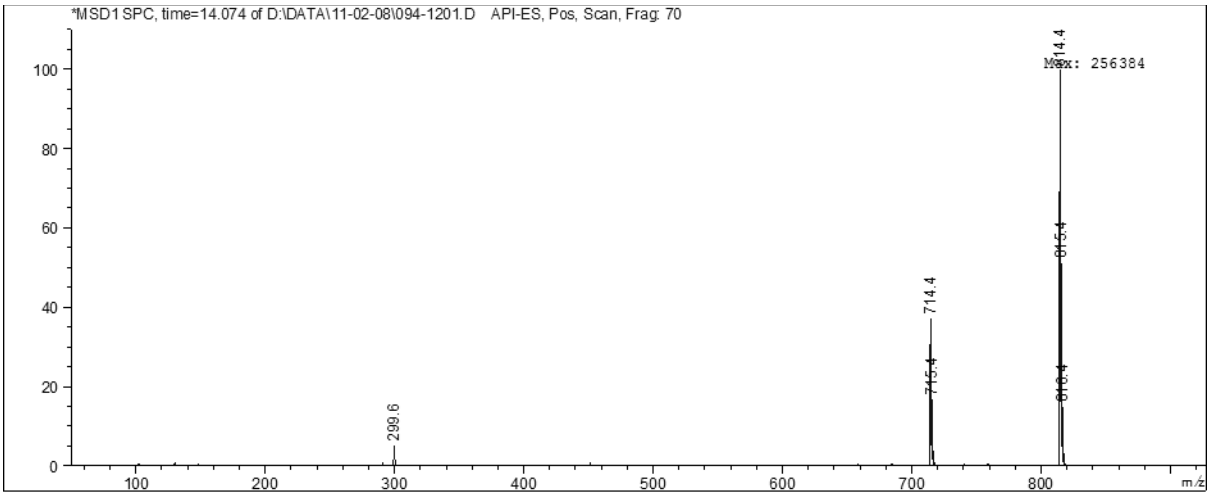
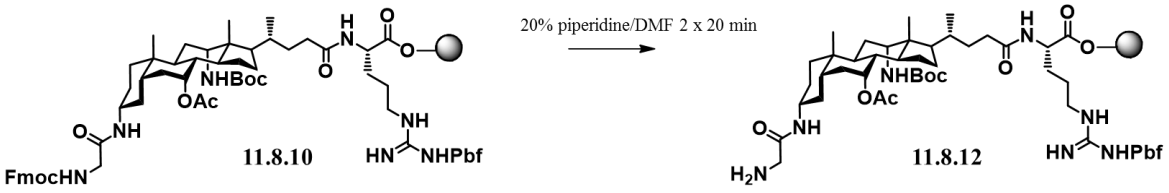


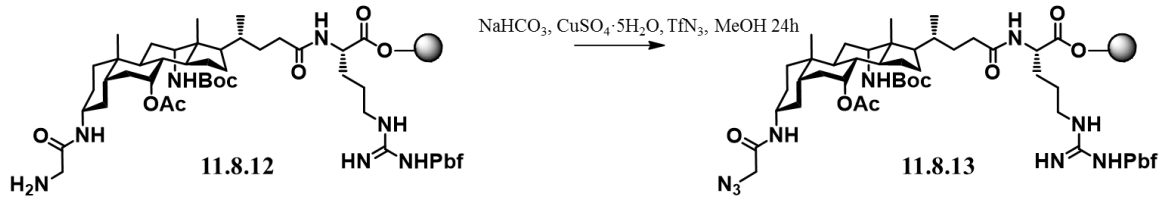
Figure 11.8.2: ESI-MS from LC-MS at t = 14.074 min for crude compound **11.8.11**. E.M calcd. for $C_{45}H_{71}N_3O_{10}$ = 813.51 and mass found 814.4.

Fmoc deprotection



Resin **11.8.10** (0.22 g, 0.55 mmol/g) was fmoc deprotected with 20% piperidine in NMP 2 x 20 min and then it was subjected to diazotransfer reaction.

Diazotransfer reaction



To the resin **11.8.2** (0.22 g, 0.55 mmol/g) were added sodium bicarbonate (0.041 g, 0.49 mmol), copper (II) sulfate pentahydrate (1 mg, 4 µmol), freshly prepared trifyl azide (2 mL) and MeOH (2 mL). The mixture was shaken at room temperature for 24 h. A small amount of resin **11.8.13** was cleaved in mild condition for LC-MS analysis. t_R 14.91 min; exact mass calculated: 1039.58; found m/z 1040.50 $[M+H]^+$ and 470.8 $[M\text{-}Boc+2H^+]/2$.

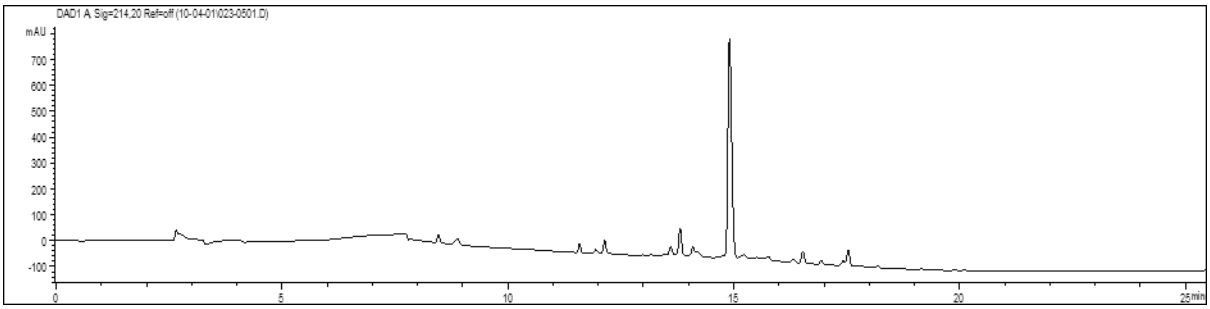


Figure 11.8.3: HPLC trace from LC-MS of compound **11.8.13** (0-100% ACN in 15 min on Luna C18 250 x 4.6 mm, 5 µm at 35°C).

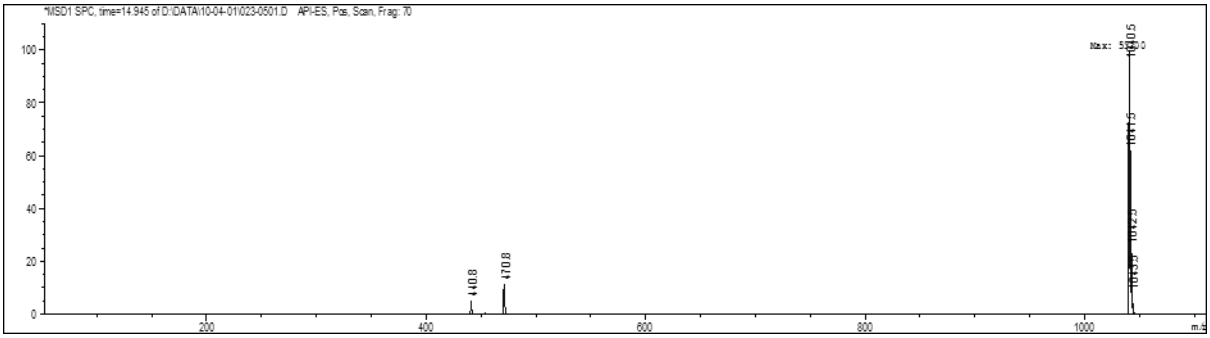
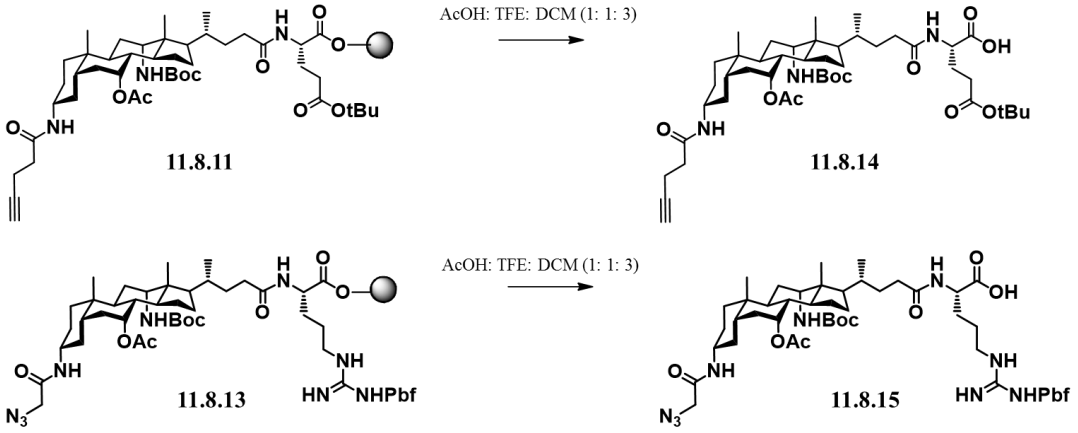


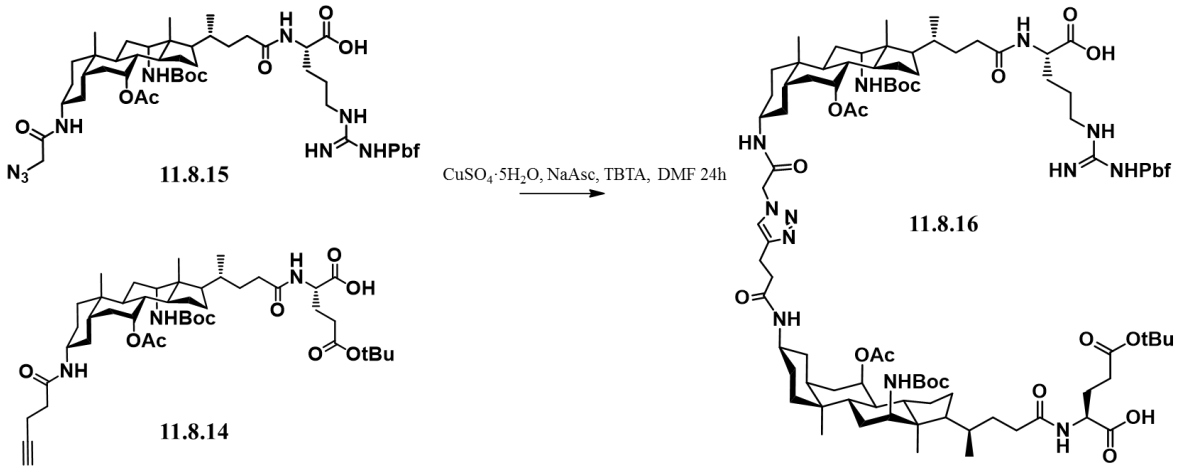
Figure 11.8.4: ESI-MS from LC-MS at t = 14.074 min for crude compound **11.8.13**. E.M calcd. for $C_{52}H_{81}N_9O_{11}S$ = 1039,58 and mass found 1040.5 $[M+H]^+$ and 470.8 $[M\text{-}Boc+2H^+]/2$.

Cleavage of alkyne and azide-containing resins



Resins **11.8.11** and **11.8.13** were treated three times with AcOH: TFE: DCM (1: 1: 3) at room temperature for 2 h and lyophilized obtaining yellowish powders **11.8.14** and **11.8.15** (86.3% and 71.2% yield).

CuAAC conjugation



Alkyne component **11.8.14** (26.8 mg, 0.033 mmol) in 75 µL DMF was added to a cone-bottom flask. Then, $CuSO_4 \cdot 5H_2O$ (5.5 mg, 0.022 mmol) in 90 µL DMF, sodium ascorbate (6.5 mg, 0.033 mmol) in 90 µL H_2O and TBTA (23.35mg, 0.044 mmol) in 90µL DMF were added to the alkyne in solution, and the mixture was stirred for 10 min. The azide component **11.8.15** (11.1 mg, 0.011 mmol) in 75 µL DMF was finally added and the reaction mixture stirred for 24h at room temperature. Then the solvent was evaporated in vacuo. The residue was diluted in EtOAc and extracted with a saturated solution of EDTA, citric acid 1M, and twice with brine. The organic phase was dried under $MgSO_4$, filtered and evaporated.

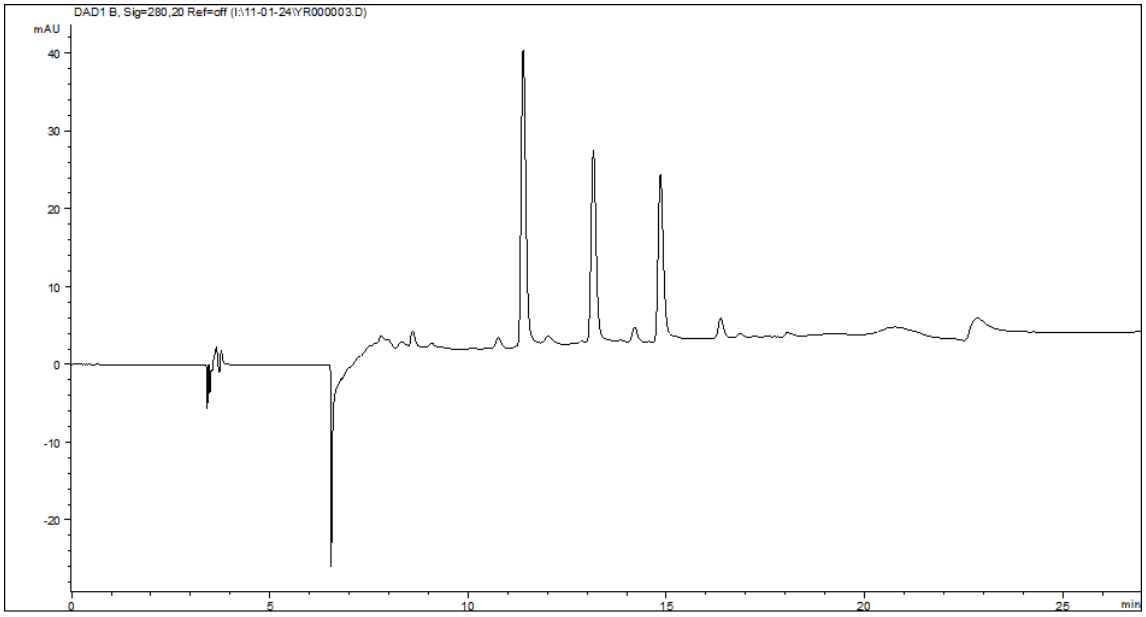


Figure 11.8.6: RP-HPLC (0-100% ACN in 15 min on Luna C18, 100 Å) of reaction mixture after 24 h. Compound **11.8.16** eluting at 14.9 min.

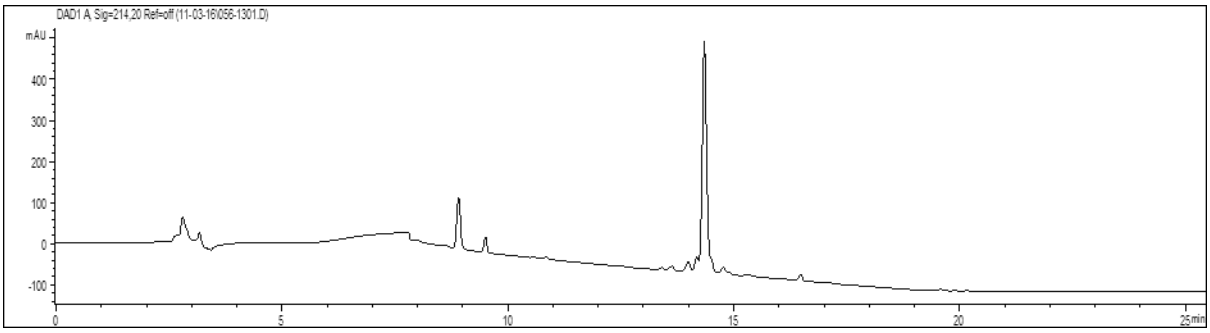


Figure 11.8.3: HPLC trace from LC-MS of pure compound **11.8.16** (0-100% ACN in 15 min on Luna C18 250 x 4.6 mm, 5 µm at 35°C).

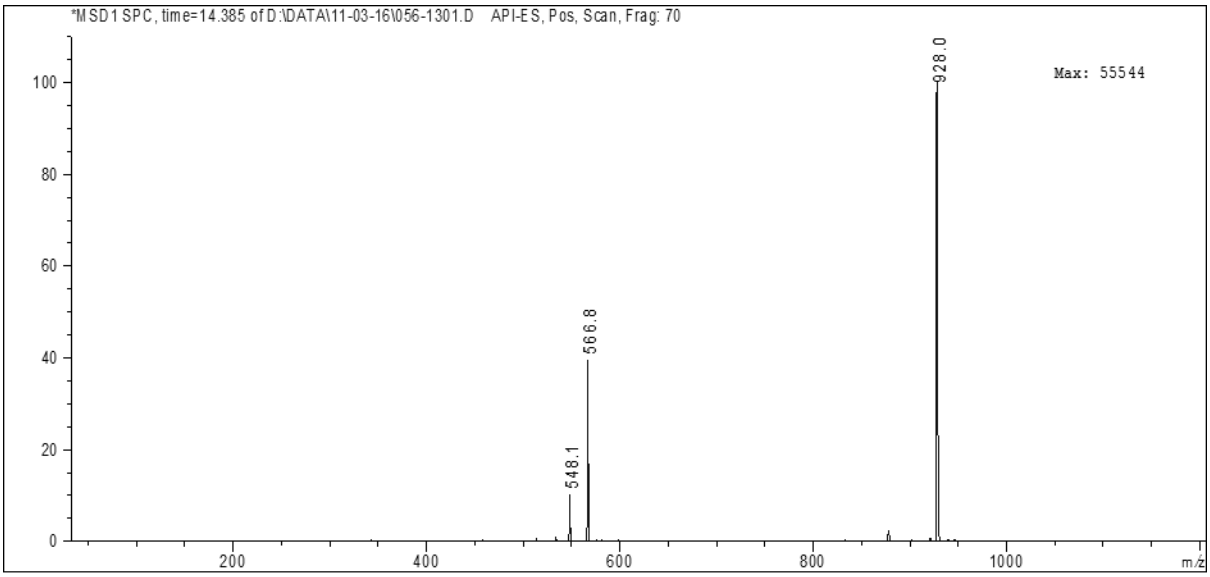


Figure 11.8.4: ESI-MS from LC-MS at t = 14.385 min for crude compound **11.8.16**. E.M calcd. for $C_{97}H_{153}N_{13}O_{20}S$ = 1853.39 and mass found 928.0 $[M/2+H]^+$.

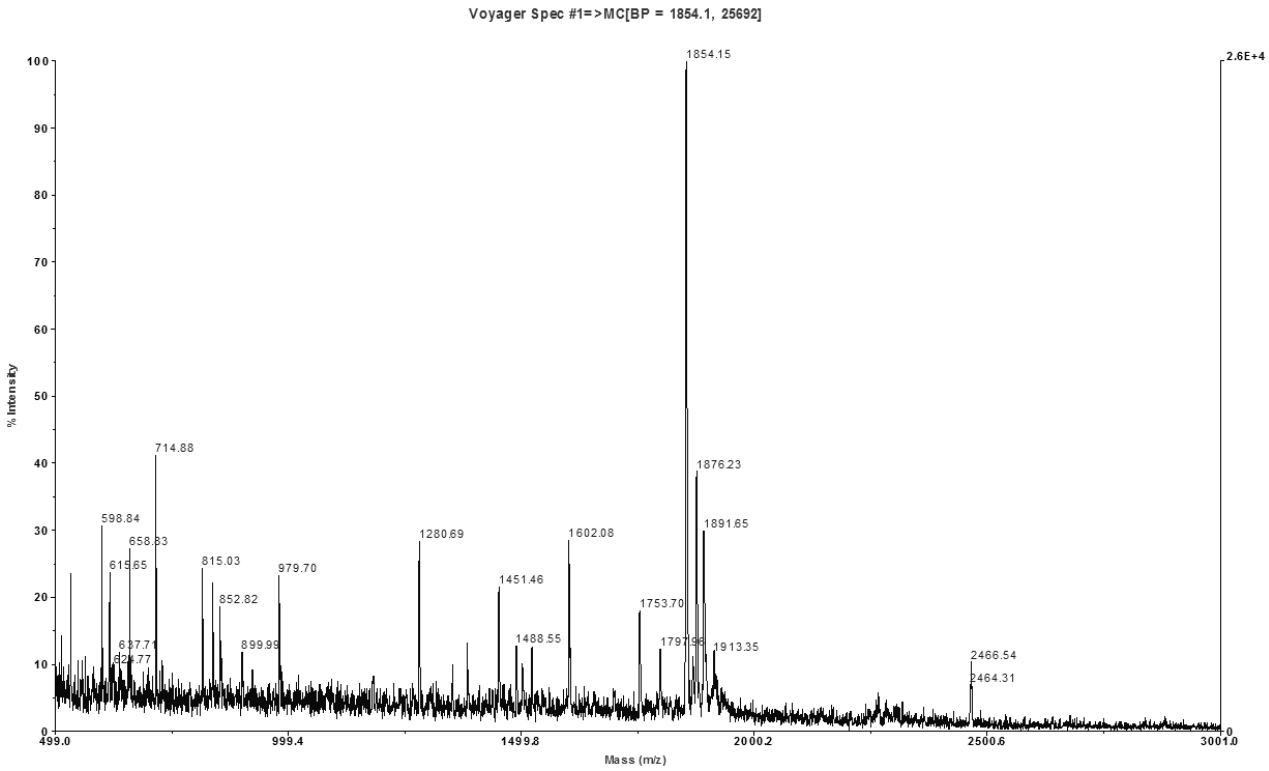
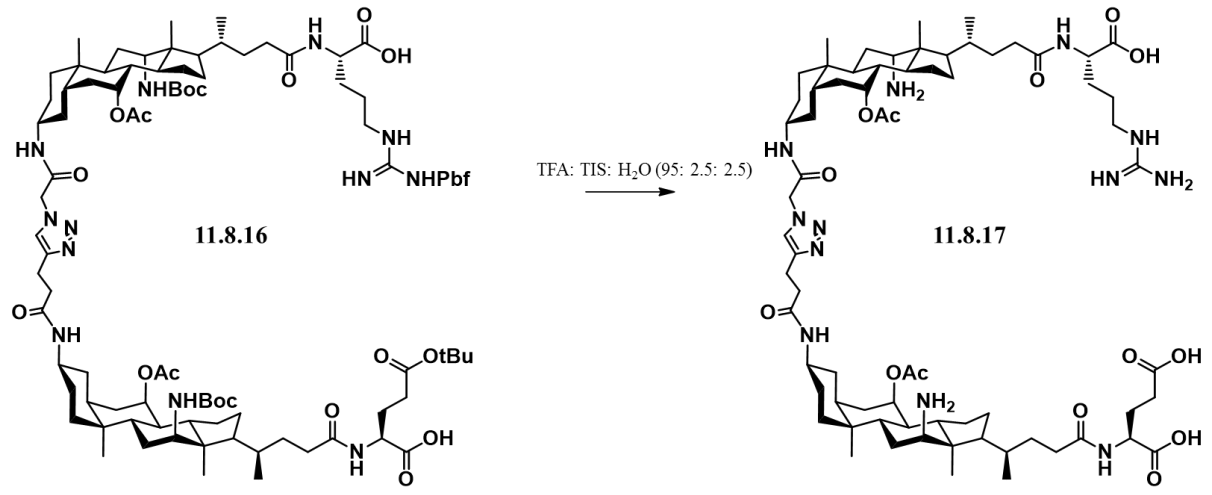


Figure 11.8.15: MALDI of crude compound **11.8.16**: calcd. EM 1852.11, MW 1853.39; Found m/z (% rel. int.) 1854.15 $[M + H]^+$.

Deprotection



The side chains were deprotected by treatment with TFA: TIS: H_2O (95: 2.5: 2.5) at room temperature for 2h. Purification was performed by RP-HPLC. The obtained material **11.8.17** was analyzed by LC-MS: t_R 11.10 min; exact mass calculated: 1344.84; found m/z 1343.80 [M-H] and 671.40 [M-2H]/2. HRMS (ESI) for $C_{70}H_{114}N_{13}O_{13}^+$: calculated 1344.8654, found 1345.8474.

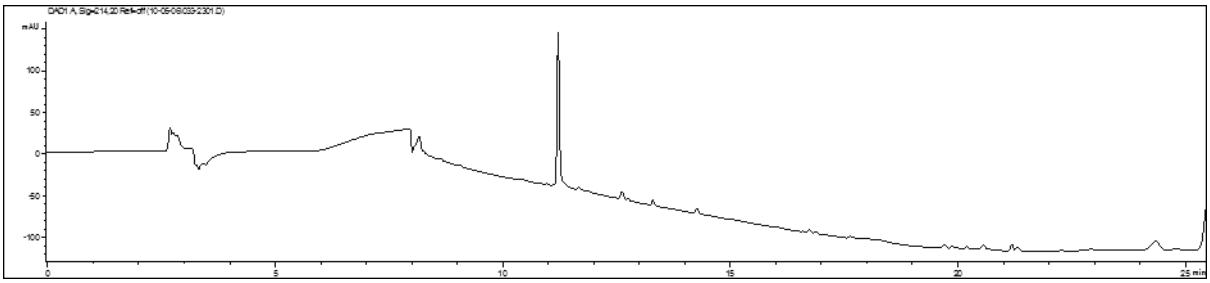


Figure 11.8.3: HPLC trace from LC-MS of pure compound **11.8.17** (0-100% ACN in 15 min on Luna C18 250 x 4.6 mm, 5 µm at 35°C).

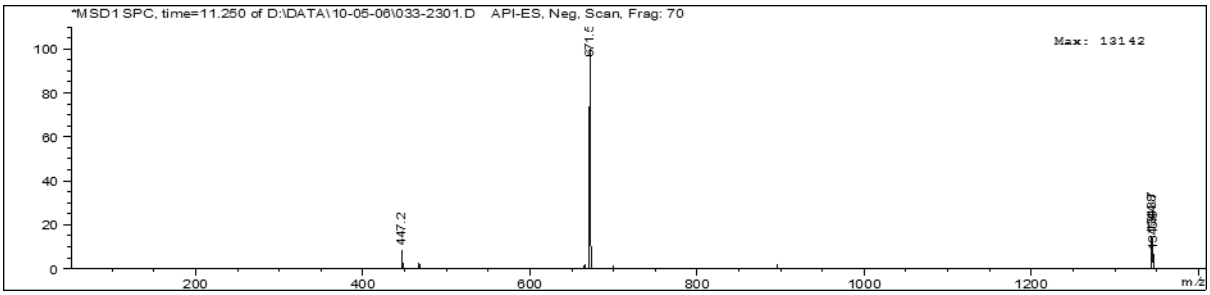


Figure 11.8.4: ESI-MS from LC-MS at t = 11.250 min for crude compound **11.8.17**. E.M calcd. for $C_{70}H_{112}N_{12}O_{14}$ = 1344.8 and mass found 671.2 $[M/2+H]^+$.

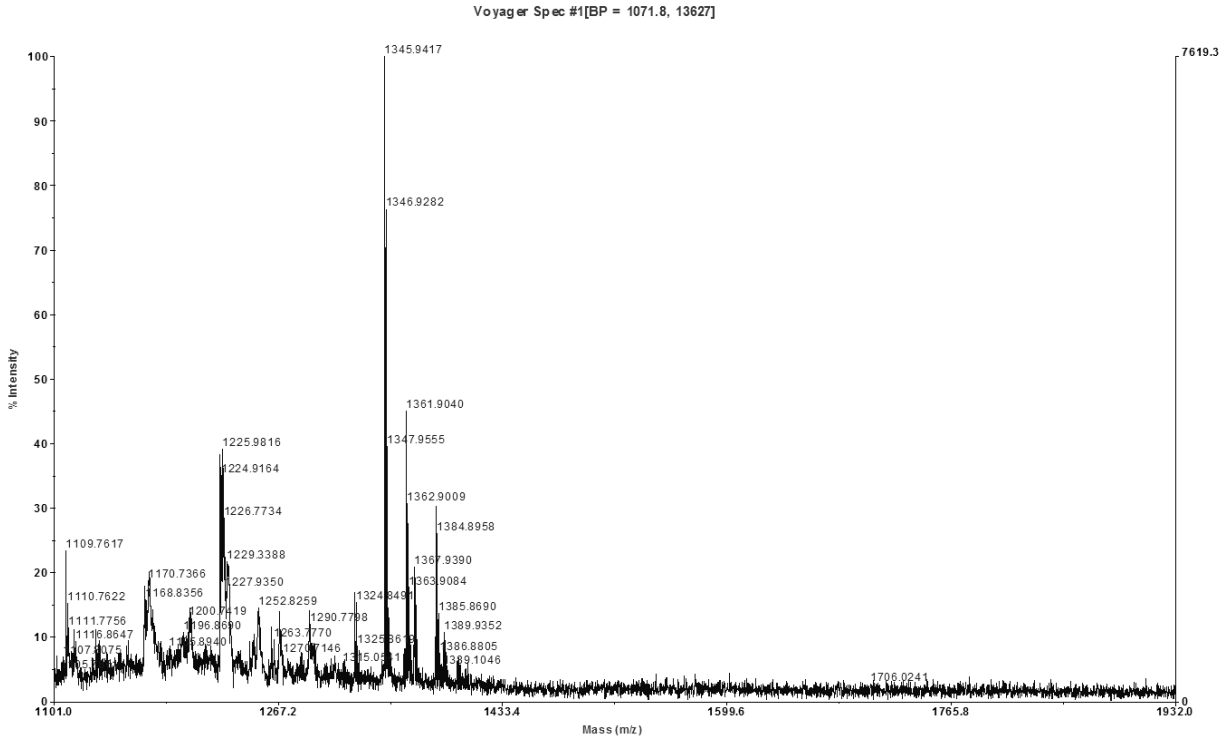
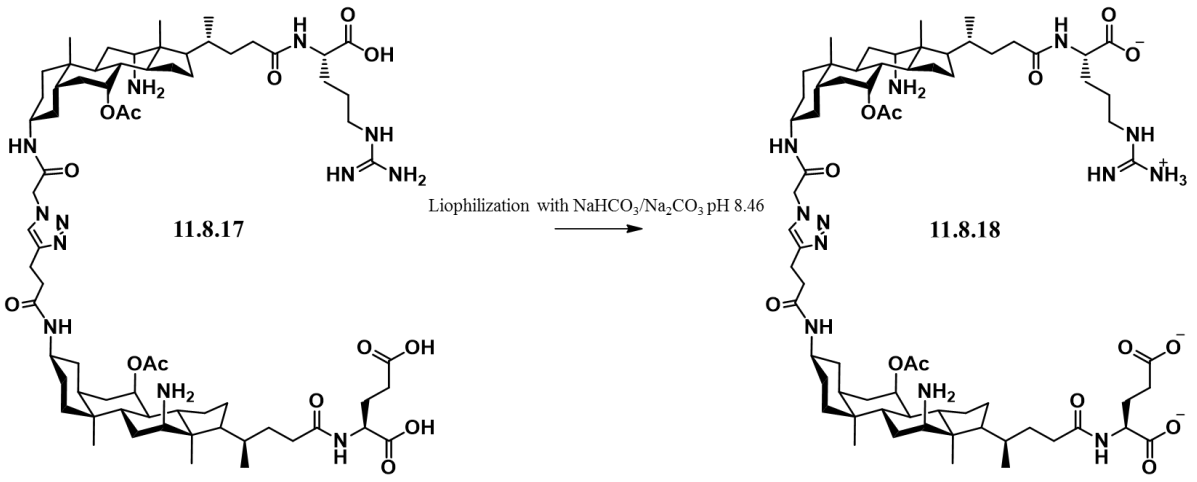


Figure 11.8.15: MALDI of crude compound **11.8.17**: calcd. EM 1344.8, MW 1345.7; Found m/z (% rel. int.) 1345.94 $[M + H]^+$.



Finally the product **11.8.17** was lyophilized overnight with $NaHCO_3/Na_2CO_3$ to obtain the product **11.8.18**. This lyophilization guaranteed the formation of the correctly charged product for NMR studies.

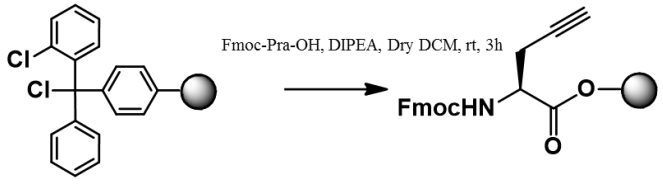
11.8.2 Development of water-soluble NHC-Cu Catalysts: Applications in Click Chemistry, Bioconjugation and Mechanistic Analysis

General information

MALDI-TOF-MS spectra were acquired with a high performance nitrogen laser (337 nm), using the positive and reflector mode with delayed extraction. All measurements were calibrated against MePEOH (Mn ≈ 2000, PD = 1.06), spotted from a MeOH (2 mg/mL) solution. Following matrix solutions were utilized (made in microtubes, stored in freezer, carefully defrosted and homogenized upon use): DHB: 2,5-dihydroxybenzoic acid (98.0 % pure, 10 mg) + CH_3CN (500 µL) + H_2O (470 µL) + TFA_{aq} (30 µL, 3 %); α-CHCA: α-cyano-4-hydroxycinnamic acid (99 % pure, 10 mg) + CH_3CN (500 µL) + H_2O (400 µL) + TFA_{aq} (100 µL, 3 %).

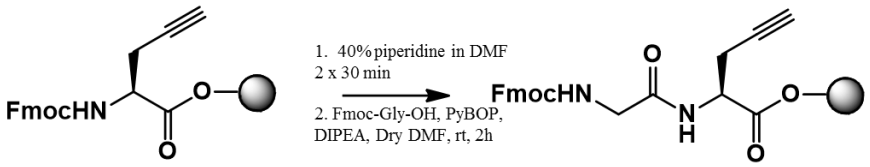
Peptide synthesis:

Immobilization of Fmoc-Pra-OH on 2-chlorotrityl chloride resin



To a suspension of 2-chlorotrutyl chloride resin **11.8.1** (0.150 g, 1.55 mmol/g) in dry DCM (10 mL/g resin), were added Fmoc-Pra-OH (Fmoc-Propargylglycine) (48 mg, 0.288 mmol) and DIPEA (0.1 mL, 0.576 mmol). The mixture was shaken at room temperature for 3 h. After the reaction, the resin was washed with DCM and Et_2O. Fmoc determination was performed to calculate the yield. The loading was calculated to be 0.56 mmol/g, indicating a coupling yield of 90%. The resin was then capped with MeOH/DIPEA in DCM (3 mL) 3 times for 2 min.

Fmoc deprotection and coupling of Fmoc-Gly-OH:

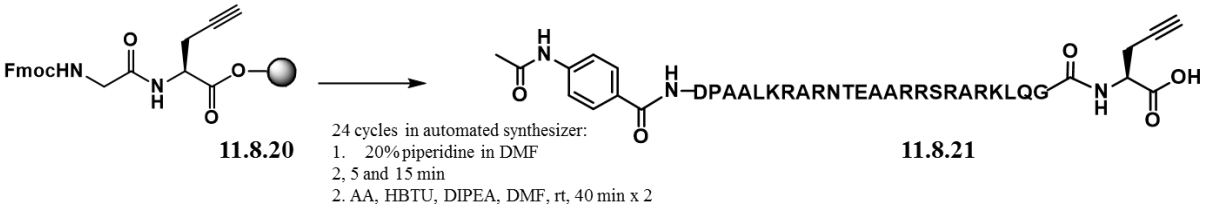


11.8.19 **11.8.20**

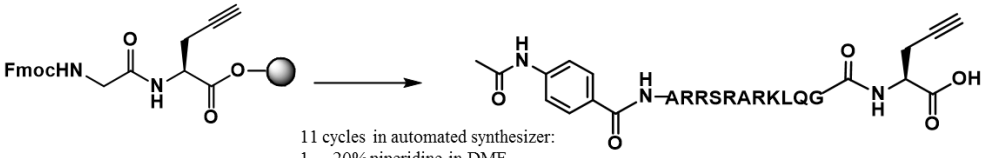
NHFmoc deprotection of compound **11.8.19**. After an initial DMF washing step, resin **11.8.19** (0.125 mmol) was successively treated twice for 30 min with a piperidine solution in DMF (40 % v/v, 3 mL) at ambient temperature, applying intermediate filtration under reduced pressure and washing with DMF, while the final resin is additionally washed with CH_3CN and DCM.

After Fmoc deprotection, Fmoc-Gly-OH (0.150 g, 0.502 mmol, 0.5 M), PyBOP (0.26 g, 0.502 mmol, 0.5 M) and DIPEA (0.087 mL, 0.502 mmol, 2 M) were added to a suspension of resin in dry DMF (3 mL). The mixture was shaken at room temperature for 3 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF successively.

Automated synthesis of linear peptide



11.8.20 **11.8.21**



11.8.20 **11.8.22**

Before automated peptide synthesis, resin **11.8.20** was manually NHFmoc-deprotected using the procedure described above.

Fmoc deprotected resin (0.125 mmol) was subjected to automated synthesis where solutions of Fmoc-N_{σ}-protected amino acids (0.5 M in NMP) were prepared. Each coupling reaction was repeated twice for 40 min (amino acids 0.625 mmol; HBTU 0.625 mmol in DMF, 0.5 M; DIPEA 0.625 mmol 2 M) and followed by Fmoc deprotection with 20% piperidine in NMP. The N-terminus was capped manually with 4-acetamidobenzoic acid (1.25 mmol, 0.5 M), PyBOP (1.25 mmol, 0.5 M), DIPEA (2.50 mmol, 2 M) in dry DMF. The peptide was then cleaved from the resin and deprotected with a cocktail of TFA:TIS:water (95: 2.5: 2.5). After precipitation in cold ether, the peptides was analyzed by RP-HPLC and MALDI.

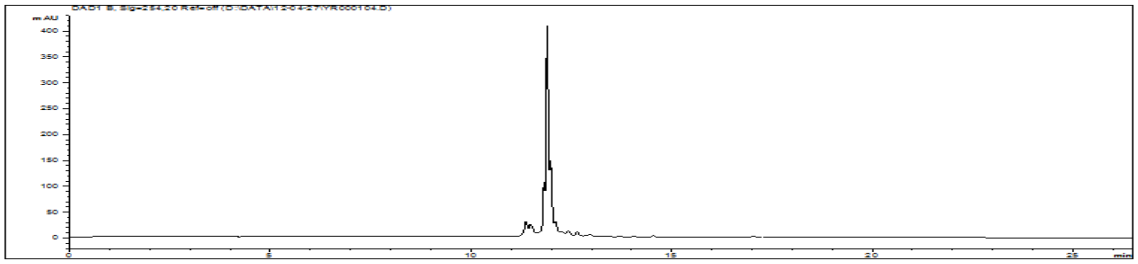


Figure 11.8.6: RP-HPLC Chromatogram of crude compound **11.8.22** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

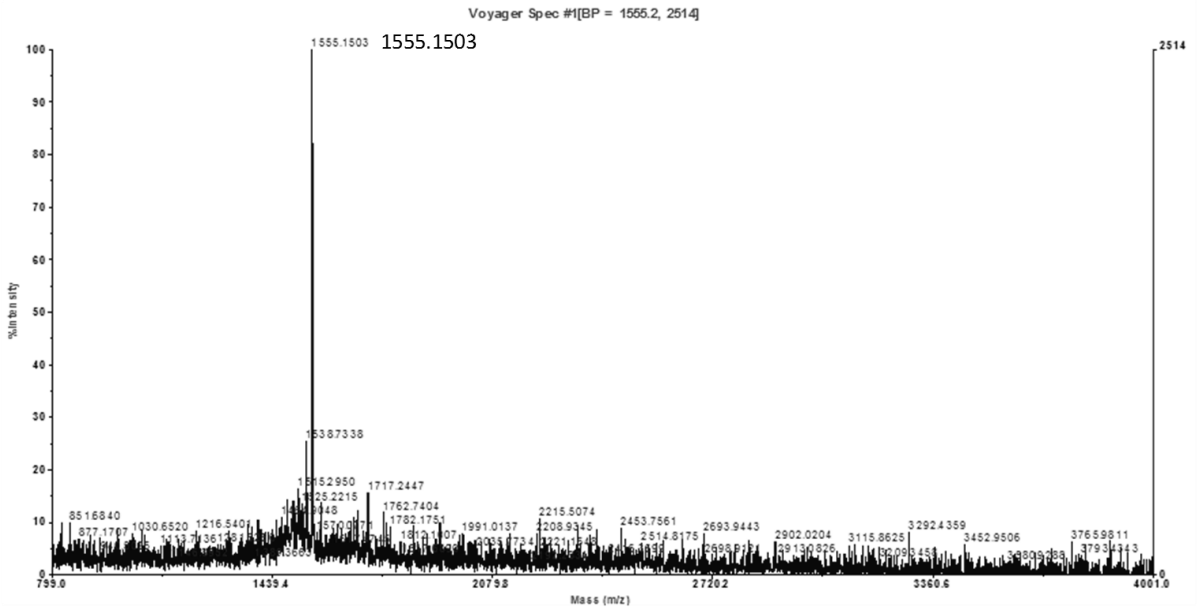


Figure 11.8.7: MALDI-TOF Spectrum of crude compound **11.8.22**. M = 1584.88. Found $[M+H]^+$ = 1585.15.

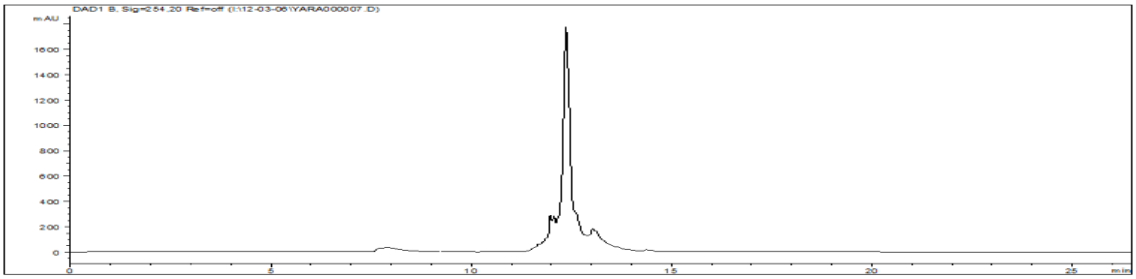


Figure 11.8.8: RP-HPLC Chromatogram of crude compound **11.8.21**. (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

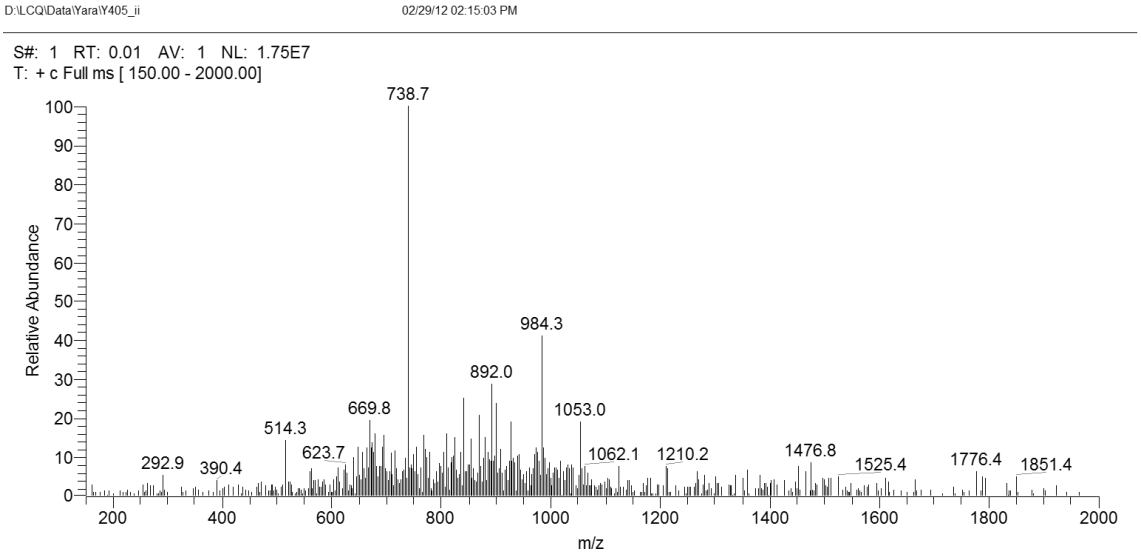
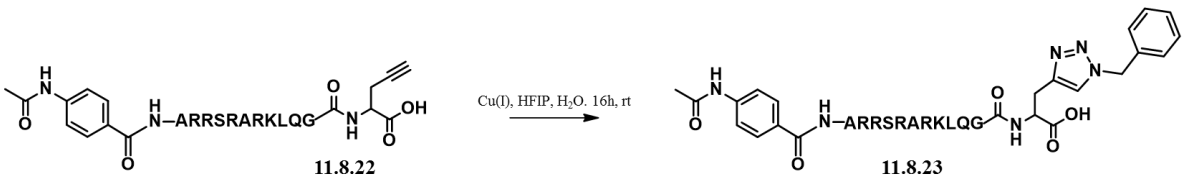


Figure 11.8.9: MS(ESI) Spectrum of crude compound **11.8.21**. M = 2949.29. Found $[M+4H]^+/4$ = 738.7; $[M+3H]^+/3$ = 984.3; $[M+2H]^+/2$ = 1476.8.

Bioconjugation via CuAAC



12mer GCN4 peptide bearing an alkyne functionality **11.8.22** ($3.86 \cdot 10^{-3}$ mmol) was dissolved in 1 mL deoxygenated water. The solution was transferred to a reaction vial under argon and the catalyst ($4.006 \cdot 10^{-3}$ mmol) was added under stirring. Then, a solution of benzylazide in DCM (0.5 M, $2.12 \cdot 10^{-2}$ mmol) was added. An extra 2.26 mL of deoxygenated water was added followed by 0.670 mL of HFIP. The reaction mixture is stirred overnight at room temperature. Analysis and purification of the final peptide are done on reverse-phase C4 HPLC.

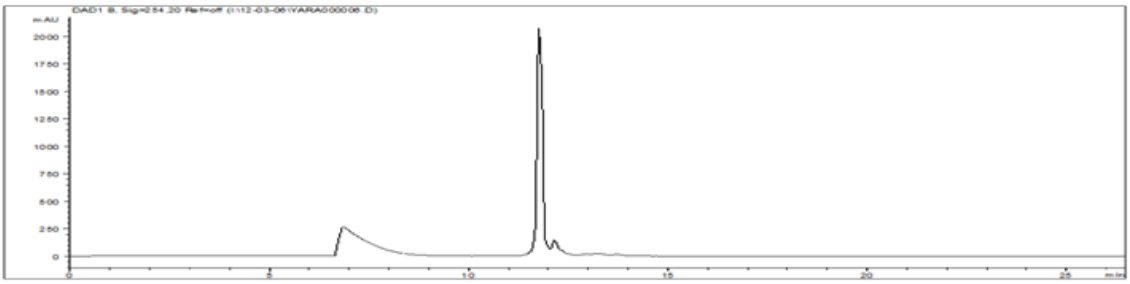


Figure 11.8.10: RP-HPLC Chromatogram of compound **11.8.23**. (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

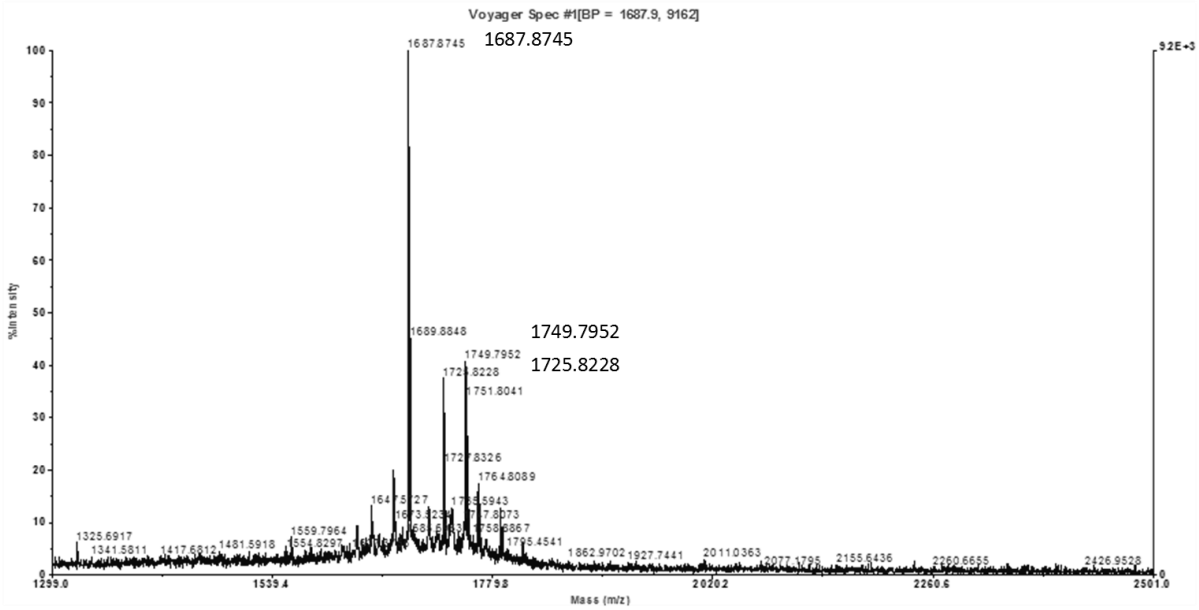
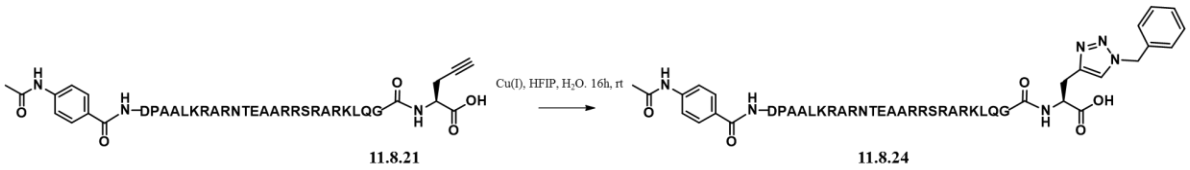


Figure 11.8.11. MALDI-TOF Spectrum of compound **11.8.23.** M = 1686.93. Found $[M+H]^+$ = 1687.8; $[M+K]^+$ = 1725.8; [M+Cu] = 1749.7.



23mer GCN4 peptide bearing an alkyne functionality **11.8.21** ($8.477 \cdot 10^{-4}$ mmol) was dissolved in 0.250 mL deoxygenated water. The solution was transferred to a reaction vial under argon and the catalyst ($9.325 \cdot 10^{-4}$ mmol) was added under stirring. Then, a solution of benzylazide in DCM (0.5 M, $4.66 \cdot 10^{-3}$ mmol) was added. An extra 0.250 mL of deoxygenated water was added followed by 50 µL of HFIP. The reaction mixture is stirred overnight at room temperature. Analysis and purification of the final peptide are done on reverse-phase C18 HPLC.

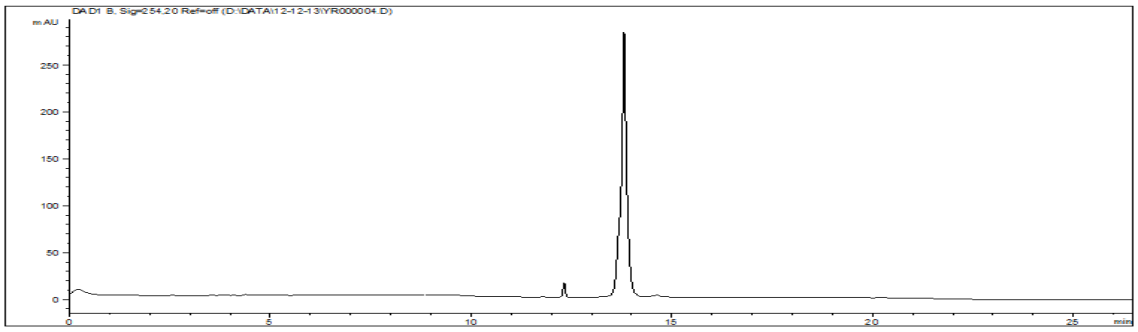


Figure 11.8.12: RP-HPLC Chromatogram of compound **11.8.24**. (C4, 300Å column using a gradient from 0 to 100 % ACN in 15 minutes).

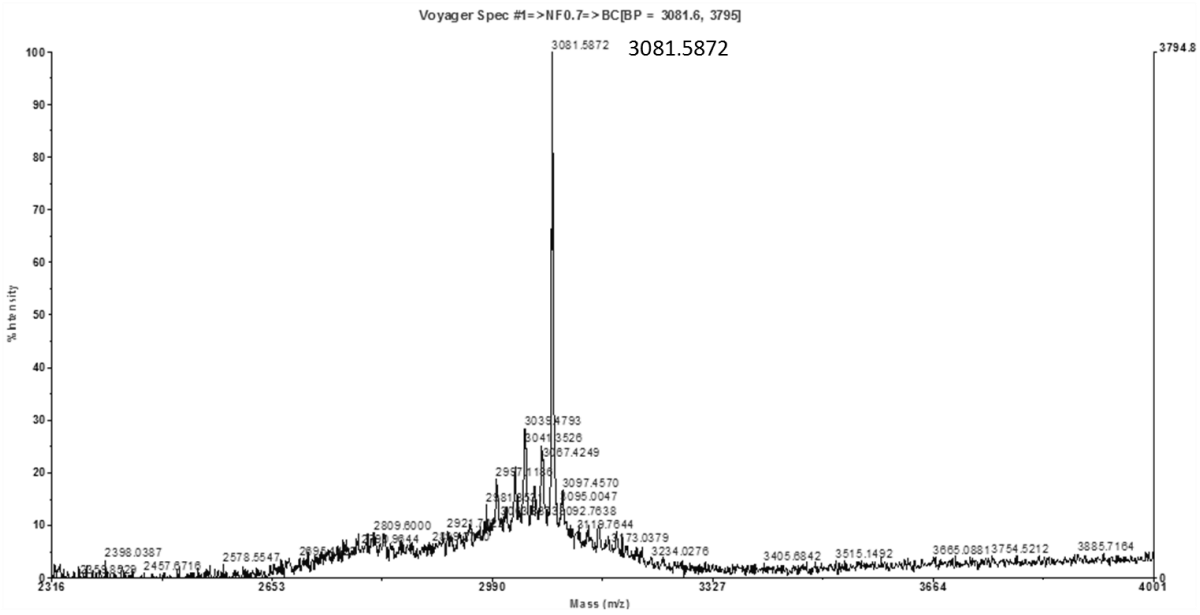
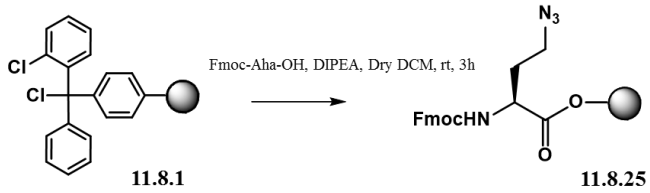


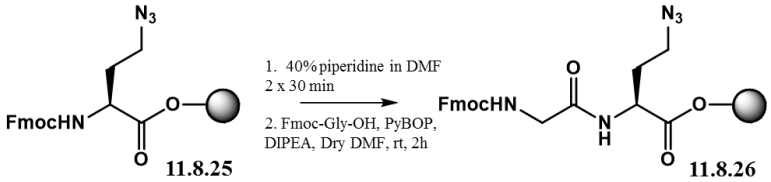
Figure 11.8.13: MALDI-TOF Spectrum of compound **11.8.24**. M = 3080.67. Found $[M+H]^+$ = 3081.58.

Immobilization of Fmoc-Aha-OH on 2-chlorotrityl resin



To a suspension of resin **11.8.1** (0.150 g, 1.55 mmol/g) in dry DCM (10 mL/g resin), were added Fmoc-Aha-OH (Fmoc-azidohomoalanine) (51 mg, 0.140 mmol) and DIPEA (0.050 mL, 0.279 mmol). The mixture was shaken at room temperature for 3 h. The loading was calculated to be 0.65 mmol/g, indicating a coupling yield of 91%. The resin was then capped with MeOH/DIPEA in DCM 3 times for 2 min.

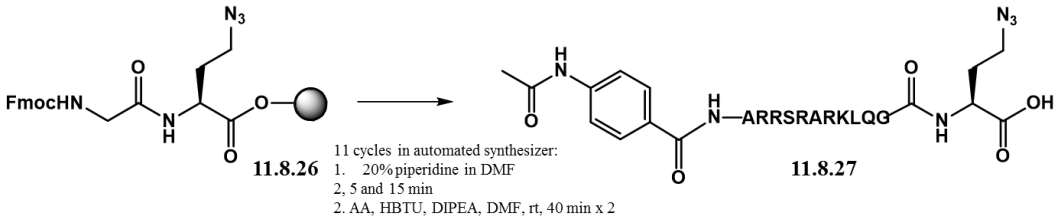
Fmoc deprotection and coupling of Fmoc-Gly-OH:



NHFmoc deprotection of construct **11.8.25**. After an initial DMF washing step, resin **11.8.25** (0.127 mmol) was successively treated twice for 30 min with a piperidine solution in DMF (40 % v/v, 3 mL) at ambient temperature, applying intermediate filtration under reduced pressure and washing with DMF, while the final resin is additionally washed with CH_3CN and DCM.

Fmoc deprotected resin (0.127 mmol). Fmoc-Gly-OH (0.151 g, 0.508 mmol, 0.5 M), PyBOP (0.264 mg, 0.508 mmol, 0.5 M) and DIPEA (0.088 mL, 0.508 mmol, 2 M) were added to a suspension of resin in dry DMF (3 mL). The mixture was shaken at room temperature for 3 h.

Automated synthesis of linear peptide



Before automated peptide synthesis, the resin **11.8.26** was manually NHFmoc-deprotected using the procedure described in.

The resin **11.8.26** (0.127 mmol) was subjected to automated synthesis where solutions of Fmoc-N_σ-protected amino acids (0.5 M in NMP) were prepared. Each reaction coupling was repeated twice for 40

min long (amino acids 0.635 mmol; HBTU 0.635 mmol in DMF, 0.5 M; DIPEA 0.635 mmol 2 M) and followed by Fmoc deprotection with 20% piperidine in NMP. The N-terminus was capped manually with 4-acetamidobenzoic acid (1.27 mmol), PyBOP (1.27 mmol), DIPEA (2.54 mmol) in dry DMF (3 mL) for 4 hours. The peptide was then cleaved for the resin and deprotected with a cocktail of TFA:TIS:water (95: 2.5: 2.5). After precipitation in cold ether, the peptide was analyzed by LC-MS(ESI) and MALDI.

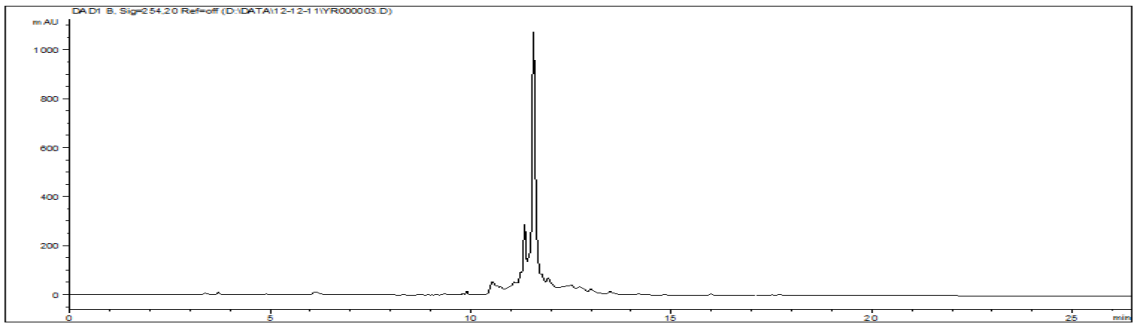


Figure 11.8.14: RP-HPLC Chromatogram of crude compound **11.8.27**. (C4, 300 Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes).

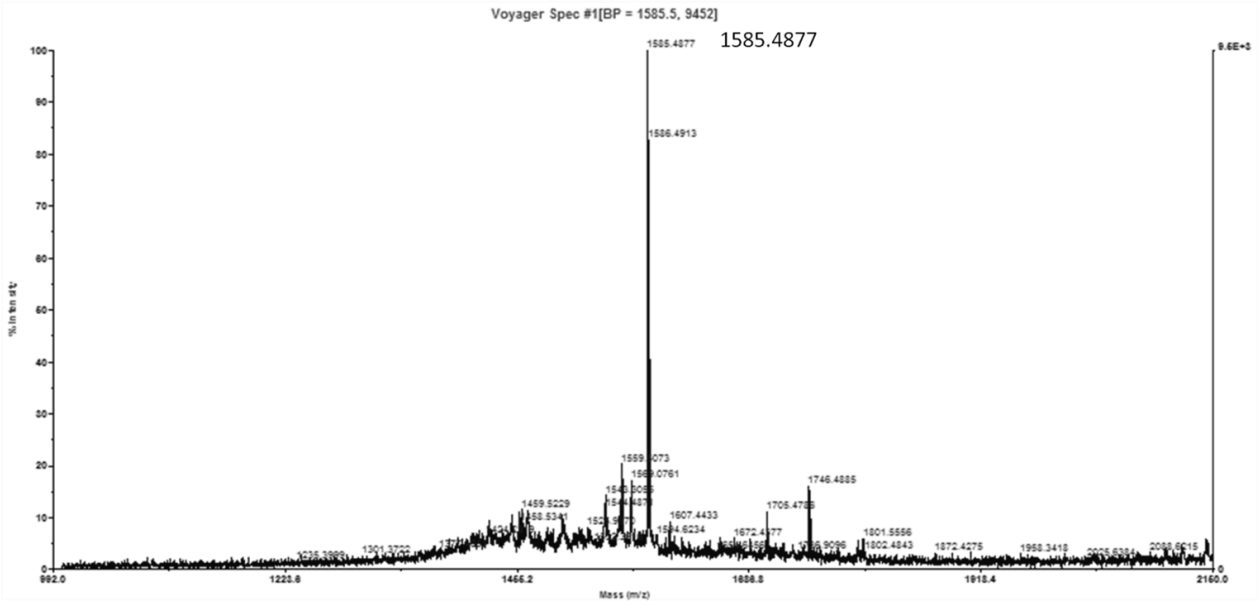
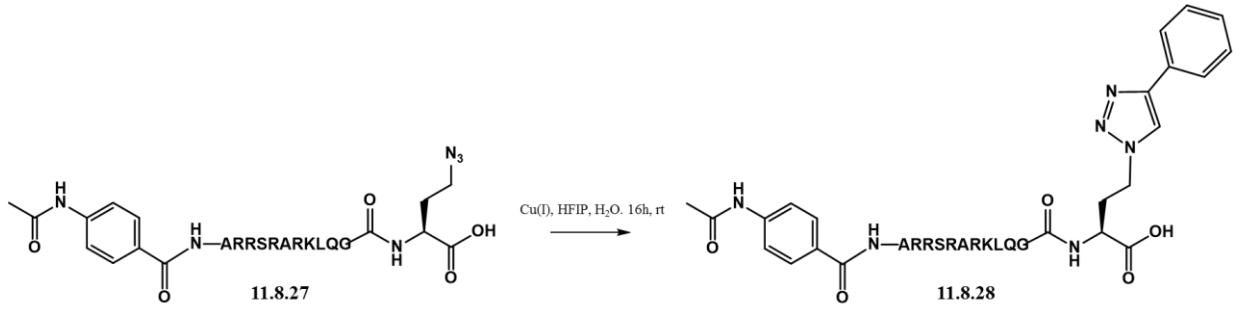


Figure 11.8.15: MALDI-TOF Spectrum of crude compound **11.8.27**. M = 1584.88. Found $[M+H]^+$ = 1585.48.

Bioconjugation via CuAAC



12mer GCN4 peptide bearing an azide functionality **11.8.27** ($3.86 \cdot 10^{-3}$ mmol) was dissolved in 1 mL deoxygenated water. The solution was transferred to a reaction vial under argon and the catalyst ($4.006 \cdot 10^{-3}$ mmol) was added under stirring. Then, a solution of benzylazide in DCM (0.5 M, $2.12 \cdot 10^{-2}$ mmol) was added. An extra 2.26 mL of deoxygenated water was added followed by 0.670 mL of HFIP. The reaction mixture is stirred overnight at room temperature. Analysis and purification of the final peptide are done on reverse-phase C4 HPLC.

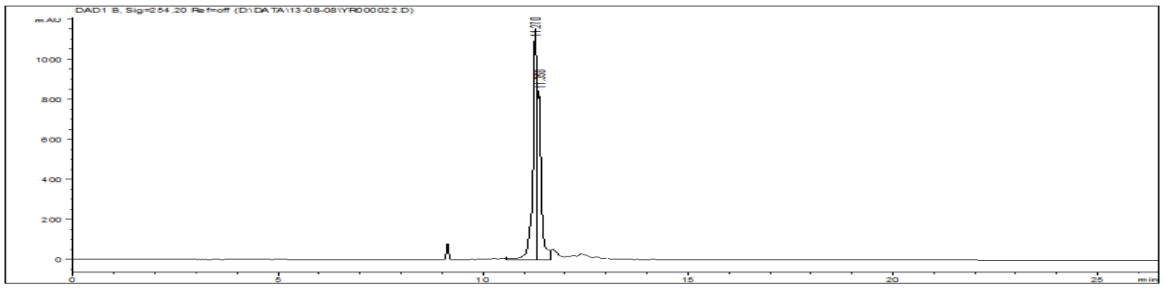


Figure 11.8.16: RP-HPLC Chromatogram of crude compound **11.8.28**. (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes). Presence of compound **11.8.27** at 11.270 min shows incompletion of the reaction.

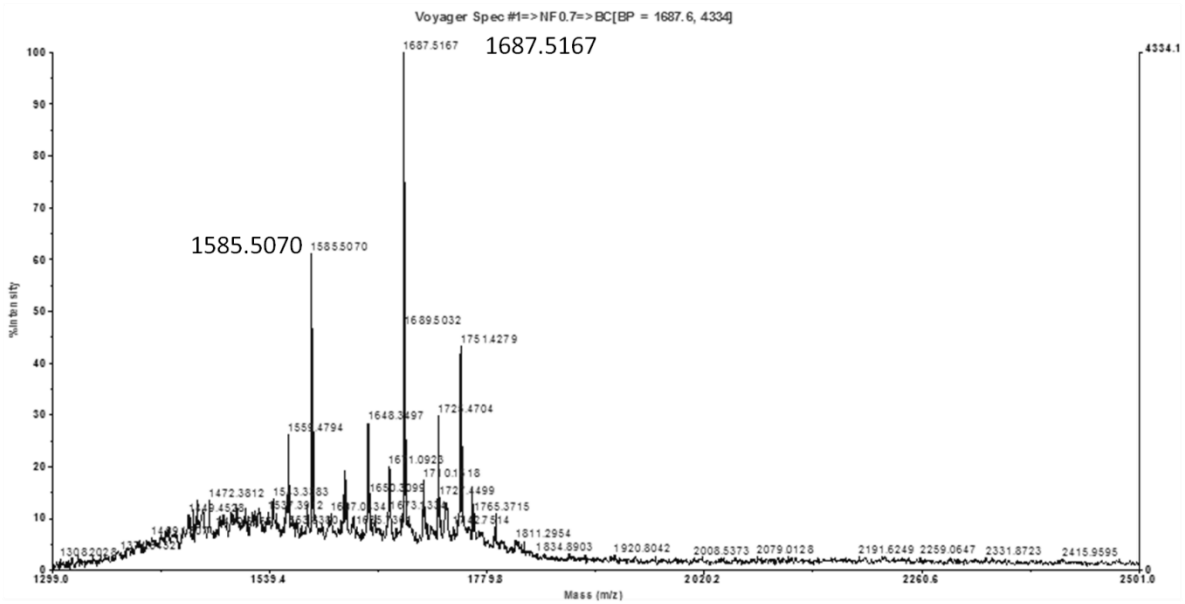


Figure 11.8.17: MALDI-TOF Spectrum of compound **11.8.28**. M = 1686.93. Found $[M+H]^+$ = 1687.51. Presence of starting compound **11.8.27.** M = 1584.88. Found $[M+H]^+$ = 1585.51.

Bibliography

(1) Miescher, F. No Title. *Hoppe-Seyler's Med. Chem. Unt* **1871**, *13*, 441–460.

(2) Austin, C. P. The completed human genome: implications for chemical biology. *Curr. Opin. Chem. Biol.* **2003**, *7*, 511–515.

(3) Jeffery, D. A.; Bogyo, M. Chemical proteomics and its application to drug discovery. *Curr. Opin. Biotechnol.* **2003**, *14*, 87–95.

(4) Baca, M.; Muir, T. W.; Schnoelzer, M.; Kent, S. B. H. Chemical Ligation of Cysteine-Containing Peptides: Synthesis of a 22 kDa Tethered Dimer of HIV-1 Protease. *J. Am. Chem. Soc.* **1995**, *117*, 1881–1887.

(5) Taylor, A. E.; Miller, C. W. On the estimation of phophorus in biological material. *J. Biol. Chem.* **1914**, *18*, 215–224.

(6) Fadok, V. A.; de Cathelineau A; Daleke, D. L.; Henson, P. M.; Bratton, D. L. Loss of phospholipid asymmetry and surface exposure of phosphatidylserine is required for phagocytosis of apoptotic cells by macrophages and fibroblasts. *J. Biol. Chem.* **2001**, *276*, 1071–1077.

(7) Davies, S. S.; Pontsler, a V; Marathe, G. K.; Harrison, K. a; Murphy, R. C.; Hinshaw, J. C.; Prestwich, G. D.; Hilaire, a S.; Prescott, S. M.; Zimmerman, G. a; et al. Oxidized alkyl phospholipids are specific, high affinity peroxisome proliferator-activated receptor gamma ligands and agonists. *J. Biol. Chem.* **2001**, *276*, 16015–16023.

(8) Daleke, D. L.; Lyles, J. V. Identification and purification of aminophospholipid flippases. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **2000**, *1486*, 108–127.

(9) Blankenberg, F. G.; Norfray, J. F. Multimodality molecular imaging of apoptosis in oncology. *AJR. Am. J. Roentgenol.* **2011**, *197*, 308–317.

(10) Zheng, H.; Wang, F.; Wang, Q.; Gao, J. Cofactor-free detection of phosphatidylserine with cyclic peptides mimicking lactadherin. *J. Am. Chem. Soc.* **2011**, *133*, 15280–15283.

(11) Haberkorn, U.; Markert, A.; Mier, W.; Askoxylakis, V.; Altmann, A. Molecular imaging of tumor metabolism and apoptosis. *Oncogene* **2011**, *30*, 4141–4151.

(12) Smith, B. A.; Smith, B. D. Biomarkers and molecular probes for cell death imaging and targeted therapeutics. *Bioconjug. Chem.* **2012**, *23*, 1989–2006.

(13) Eisenhut, M.; Haberkorn, U. Molecular Position of Radiolabels and Its Impact on Functional Integrity of Proteins. *J. Nucl. Med.* **2006**, *47*, 1400–1402.

(14) Annexin V molecular structure, ligand binding and biological function. In *Annexins: Molecular Structure to Cellular Function*; Seaton, B. A., Ed.; R.G. Landes Co: Austin, TX, 1996; pp. 15–29.

(15) Tait, J.; Gibson, D.; Fujikawa, K. Phospholipid binding properties of human placental anticoagulant protein-I, a member of the lipocortin family. *J. Biol. Chem.* **1989**, *264*, 7944–7949.

(16) Swairjo, M. A.; Concha, N. O.; Kaetzel, M. A.; Dedman, J. R.; Seaton, B. A. Ca2+-bridging mechanism and phospholipid head group recognition in the membrane-binding protein annexin V. *Nat. Struct. Biol.* **1995**, *2*, 968–974.

(17) Zweifach, A. FM1-43 reports plasma membrane phospholipid scrambling in T-lymphocytes. *Biochem. J.* **2000**, *349*, 255–260.

(18) Balasubramanian, K.; Bevers, E. M.; Willems, G. M.; Schroit, A. J. Binding of Annexin V to Membrane Products of Lipid Peroxidation †. *Biochemistry* **2001**, *40*, 8672–8676.

(19) Ojida, A.; Mito-Oka, Y.; Inoue, M.-A.; Hamachi, I. First artificial receptors and chemosensors toward phosphorylated peptide in aqueous solution. *J. Am. Chem. Soc.* **2002**, *124*, 6256–6258.

(20) Koulov, a V; Stucker, K. a; Lakshmi, C.; Robinson, J. P.; Smith, B. D. Detection of apoptotic cells using a synthetic fluorescent sensor for membrane surfaces that contain phosphatidylserine. *Cell Death Differ.* **2003**, *10*, 1357–1359.

(21) Lakshmi, C.; Hanshaw, R. G.; Smith, B. D. Fluorophore-linked zinc(II)dipicolylamine coordination complexes as sensors for phosphatidylserine-containing membranes. *Tetrahedron* **2004**, *60*, 11307–11315.

(22) Hanshaw, R. G.; Lakshmi, C.; Lambert, T. N.; Johnson, J. R.; Smith, B. D. Fluorescent detection of apoptotic cells by using zinc coordination complexes with a selective affinity for membrane surfaces enriched with phosphatidylserine. *Chembiochem* **2005**, *6*, 2214–2220.

(23) DiVittorio, K. M.; Johnson, J. R.; Johansson, E.; Reynolds, A. J.; Jolliffe, K. a; Smith, B. D. Synthetic peptides with selective affinity for apoptotic cells. *Org. Biomol. Chem.* **2006**, *4*, 1966–1976.

(24) Hanshaw, R. G.; O'Neil, E. J.; Foley, M.; Carpenter, R. T.; Smith, B. D. Indicator displacement assays that detect bilayer membranes enriched in phosphatidylserine. *J. Mater. Chem.* **2005**, *15*, 2707.

(25) Zhu, C.; Yang, Q.; Liu, L.; Wang, S. A potent fluorescent probe for the detection of cell apoptosis. *Chem. Commun. (Camb).* **2011**, *47*, 5524–5526.

(26) Smith, B. a; Akers, W. J.; Leevy, W. M.; Lampkins, A. J.; Xiao, S.; Wolter, W.; Suckow, M. a; Achilefu, S.; Smith, B. D. Optical imaging of mammary and prostate tumors in living animals using a synthetic near infrared zinc(II)-dipicolylamine probe for anionic cell surfaces. *J. Am. Chem. Soc.* **2010**, *132*, 67–69.

(27) Smith, B. a; Xiao, S.; Wolter, W.; Wheeler, J.; Suckow, M. a; Smith, B. D. In vivo targeting of cell death using a synthetic fluorescent molecular probe. *Apoptosis* **2011**, *16*, 722–731.

(28) Turkyilmaz, S.; Rice, D. R.; Palumbo, R.; Smith, B. D. Selective recognition of anionic cell membranes using targeted liposomes coated with zinc(ii)-bis(dipicolylamine) affinity units. *Org. Biomol. Chem.* **2014**, *12*, 5645–5655.

(29) Maeda, H. Tumor-selective delivery of macromolecular drugs via the EPR effect: background and future prospects. *Bioconjug. Chem.* **2010**, *21*, 797–802.

(30) Barenholz, Y. Doxil®--the first FDA-approved nano-drug: lessons learned. *J. Control. Release* **2012**, *160*, 117–134.

(31) Boon, J. M.; Smith, B. D. Chemical control of phospholipid distribution across bilayer membranes. *Med. Res. Rev.* **2002**, *22*, 251–281.

(32) Kato, N.; Nakanishi, M.; Hirashima, N. Transbilayer Asymmetry of Phospholipids in the Plasma Membrane Regulates Exocytotic Release in Mast Cells. *Biochemistry* **2002**, *41*, 8068–8074.

(33) Manno, S.; Takakuwa, Y.; Mohandas, N. Identification of a functional role for lipid asymmetry in biological membranes: Phosphatidylserine-skeletal protein interactions modulate membrane stability. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1943–1948.

(34) Bevers, E. M.; Comfurius, P.; Van Rijn, J. L. M. L.; Hemker, H. C. Generation of Prothrombin-Converting Activity and the Exposure of Phosphatidylserine at the Outer Surface of Platelets. *Eur. J. Biochem.* **1982**, *122*, 429–436.

(35) Lee, G.; Pollard, H. B.; Arispe, N. Annexin 5 and apolipoprotein E2 protect against Alzheimer's amyloid-β-peptide cytotoxicity by competitive inhibition at a common phosphatidylserine interaction site. *Peptides* **2002**, *23*, 1249–1263.

(36) Boon, J. M.; Lambert, T. N.; Sisson, A. L.; Davis, A. P.; Smith, B. D. Facilitated phosphatidylserine (PS) flip-flop and thrombin activation using a synthetic PS scramblase. *J. Am. Chem. Soc.* **2003**, *125*, 8195–8201.

(37) Valiyaveettil, S.; Engbersen, J. F. J.; Verboom, W.; Reinhoudt, D. N. Synthesis and Complexation Studies of Neutral Anion Receptors. *Angew. Chemie Int. Ed. English* **1993**, *32*, 900–901.

(38) Scheerder, J.; Fochi, M.; Engbersen, J. F. J.; Reinhoudt, D. N. Urea-Derivatized p-tert-Butylcalix[4]arenes: Neutral Ligands for Selective Anion Complexation. *J. Org. Chem.* **1994**, *59*, 7815–7820.

(39) Davis, A. P.; Dresen, S.; Lawless, L. J. Mitsunobu reactions with methanesulfonic acid; The replacement of equatorial hydroxyl groups by azide with net retention of configuration. *Tetrahedron Lett.* **1997**, *38*, 4305–4308.

(40) Broderick, S.; Davis, A. P.; Williams, R. P. The "triamino-analogue" of methyl cholate; a facial amphiphile and scaffold with potential for combinatorial and molecular recognition chemistry. *Tetrahedron Lett.* **1998**, *39*, 6083–6086.

(41) Davis, A. P.; Pérez-Payán, M. N. The "Triamino-analogue" of Methyl Cholate; A Practical, Large-Scale Synthesis. *Synlett* **1999**, 991–993.

(42) Lambert, T. N.; Boon, J. M.; Smith, B. D.; Pérez-Payán, M. N.; Davis, A. P. Facilitated phospholipid flip-flop using synthetic steroid-derived translocases. *J. Am. Chem. Soc.* **2002**, *124*, 5276–5277.

(43) Khatri, V. K.; Upreti, S.; Pandey, P. S. Novel bile acid-based cyclic bisimidazolium receptors for anion recognition. *Org. Lett.* **2006**, *8*, 1755–1758.

(44) Ornelas, C.; Ruiz Aranzaes, J.; Cloutet, E.; Alves, S.; Astruc, D. Click assembly of 1,2,3-triazole-linked dendrimers, including ferrocenyl dendrimers, which sense both oxo anions and metal cations. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 872–877.

(45) Palmer, M. H.; Findlay, R. H.; Gaskell, A. J. Electronic charge distribution and moments of five- and six-membered heterocycles. *J. Chem. Soc. Perkin Trans. 2* **1974**, 420.

(46) Hua, Y.; Flood, A. H. Click chemistry generates privileged CH hydrogen-bonding triazoles: the latest addition to anion supramolecular chemistry. *Chem. Soc. Rev.* **2010**, *39*, 1262.

(47) Horne, W. S.; Yadav, M. K.; Stout, C. D.; Ghadiri, M. R. Heterocyclic Peptide Backbone Modifications in an Î±-Helical Coiled Coil. *J. Am. Chem. Soc.* **2004**, *126*, 15366–15367.

(48) Verzele, D.; Madder, A. Short Synthesis of Orthogonally Protected 3α,12α-Diamino-5β-cholan-24-oic Acid, a Dipodal Steroid Scaffold for Combinatorial Chemistry. *Eur. J. Org. Chem.* **2007**, 1793–1797.

(49) Wenska, M.; Alvira, M.; Steunenberg, P.; Stenberg, A.; Murtola, M.; Stroemberg, R. An activated triple bond linker enables “click” attachment of peptides to oligonucleotides on solid support. *Nucleic Acids Res.* **2011**, *39*, 9047–9059.

(50) Carter, P. S. H. and M. *DNA Replication-Current Advances*; Seligmann, H., Ed.; InTech, 2011.

(51) Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 188–200.

(52) Arndt, K.; Fink, G. R. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5’ TGACTC 3' sequences. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 8516–8520.

(53) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted Î± helices: crystal structure of the protein-DNA complex. *Cell (Cambridge, Mass.)* **1992**, *71*, 1223–1237.

(54) Talanian, R. V; McKnight, C. J.; Kim, P. S. Sequence-specific DNA binding by a short peptide dimer. *Sci. (Washington, D. C., 1883-)* **1990**, *249*, 769–771.

(55) Cuenoud, B.; Schepartz, A. Altered specificity of DNA-binding proteins with transition metal dimerization domains. *Sci. (Washington, D. C., 1883-)* **1993**, *259*, 510–513.

(56) Ueno, M.; Murakami, A.; Makino, K.; Morii, T. Arranging quaternary structure of peptides by cyclodextrin-guest inclusion complex: sequence-specific DNA binding by a peptide dimer with artificial dimerization module. *J. Am. Chem. Soc.* **1993**, *115*, 12575–12576.

(57) Caamano, A. M.; Vazquez, M. E.; Martinez-Costas, J.; Castedo, L.; Mascarenas, J. L. A light-modulated sequence-specific DNA-binding peptide. *Angew. Chem., Int. Ed.* **2000**, *39*, 3104–3107.

(58) Carrette, L. L. G.; Morii, T.; Madder, A. Peptidosteroid Tweezers Revisited: DNA Binding Through an Optimised Design. *European J. Org. Chem.* **2014**, *2014*, 2883–2891.

(59) Savage, P. B.; Li, C.; Taotafa, U.; Ding, B.; Guan, Q. Antibacterial properties of cationic steroid antibiotics. *FEMS Microbiol. Lett.* **2002**, *217*, 1–7.

(60) Savage, P. B. Design, Synthesis and Characterization of Cationic Peptide and Steroid Antibiotics. *European J. Org. Chem.* **2002**, *2002*, 759–768.

(61) Nath, S.; Maitra, U. A simple and general strategy for the design of fluorescent cation sensor beads. *Org. Lett.* **2006**, *8*, 3239–3242.

(62) Davis, A. P.; Joos, J.-B. Steroids as organising elements in anion receptors. *Coord. Chem. Rev.* **2003**, *240*, 143–156.

(63) Ghosh, S.; Choudhury, A. R.; Guru Row, T. N.; Maitra, U. Selective and unusual fluoride ion complexation by a steroidal receptor using OH...F- and CH...F- interactions: a new motif for anion coordination? *Org. Lett.* **2005**, *7*, 1441–1444.

(64) Yoshii, M.; Yamamura, M.; Satake, A.; Kobuke, Y. Supramolecular ion channels from a transmembrane bischolic acid derivative showing two discrete conductances. *Org. Biomol. Chem.* **2004**, *2*, 2619–2623.

(65) Del Amo, V.; Siracusa, L.; Markidis, T.; Baragaña, B.; Bhattarai, K. M.; Galobardes, M.; Naredo, G.; Pérez-Payán, M. N.; Davis, A. P. Differentially-protected steroidal triamines; scaffolds with potential for medicinal, supramolecular, and combinatorial chemistry. *Org. Biomol. Chem.* **2004**, *2*, 3320–3328.

(66) Enhsen, A.; Kramer, W.; Wess, G. Bile acids in drug discovery. *Drug Discov. Today* **1998**, *3*, 409–418.

(67) Kramer, W.; Wess, G.; Enhsen, A.; Falk, E.; Hoffmann, A.; Neckermann, G.; Schubert, G.; Urmann, M. Modified bile acids as carriers for peptides and drugs. *J. Control. Release* **1997**, *46*, 17–30.

(68) Bode, C. A.; Bechet, T.; Prodhomme, E.; Gheysen, K.; Gregoir, P.; Martins, J. C.; Muller, C. P.; Madder, A. Towards the conformational mimicry of the measles virus HNE loop: design, synthesis and biological evaluation of a cyclic bile acid-peptide conjugate. *Org. Biomol. Chem.* **2009**, *7*, 3391–3399.

(69) Salunke, D. B.; Hazra, B. G.; Pore, V. S. Steroidal conjugates and their pharmacological applications. *Curr. Med. Chem.* **2006**, *13*, 813–847.

(70) Bode, C. A.; Muller, C. P.; Madder, A. Validation of a solid-phase-bound steroid scaffold for the synthesis of novel cyclic peptidosteroids. *J. Pept. Sci.* **2007**, *13*, 702–708.

(71) Madder, A.; Li, L.; De, M. H.; Farcy, N.; Van, H. D.; Fant, F.; Vanhoenacker, G.; Sandra, P.; Davis, A. P.; De, C. P. J. Evaluation of a two-stage screening procedure in the combinatorial search for serine protease-like activity. *J. Comb. Chem.* **2002**, *4*, 552–562.

(72) Li, H.; Wang, L.-X. Cholic acid as template for multivalent peptide assembly. *Org. Biomol. Chem.* **2003**, *1*, 3507–3513.

(73) Souza, L. J. D.; Maitra, U. Design , Synthesis , and Evaluation of Bile Acid-Based Molecular Tweezers. **1996**, *3263*, 9494–9502.

(74) Verzele, D.; Goeman, J. L.; Madder, A. LC-(TIC/EIC)-MS as tool in the analysis of diastereomeric 3,12-diamino-analogs of deoxycholic acid. *Ark. (Gainesville, FL, U. S.)* **2007**, 325–336.

(75) Cheng, Y.; Suenaga, T.; Still, W. C. Sequence-Selective Peptide Binding with a Peptido-A , B-trans -steroidal Receptor Selected from an Encoded Combinatorial Receptor Library. *J. Am. Chem. Soc.* **1996**, *12*, 1813–1814.

(76) Hsieh, H.-P.; Muller, J. G.; Burrows, C. J. Synthesis and DNA binding properties of C3-, C12-, and C24- substituted amino-steroids derived from bile acids. *Bioorg. Med. Chem.* **1995**, *3*, 823–838.

(77) Barry, J. F.; Davis, A. P.; Nieves Pérez-Payan, M.; Elsegood, M. R. J.; Jackson, R. F. W.; Gennari, C.; Piarulli, U.; Gude, M. A trifunctional steroid-based scaffold for combinatorial chemistry. *Tetrahedron Lett.* **1999**, *40*, 2849–2852.

(78) Verzele, D. Development of Miniature b(-HLH-)ZIP Peptidosteroid Models. *PhD Thesis Dev. Miniat. b(-HLH-)ZIP Pept. Model.* **2010**.

(79) Verzele, D.; Madder, A. Synthetic Progress in cMyc-Max Oncoprotein Miniaturization: Semi-Online Monitoring Gives Solid-Phase Access to Hydrophobic b(-HLH-)ZIP Peptidosteroid Tweezers. *European J. Org. Chem.* **2013**, *2013*, 673–687.

(80) Carrette, L. L. G. Peptide-DNA interactions: a study using protein models and the furan-oxidation cross-link methodology. *PhD Thesis Pept. Interact. a study using protein Model. furan-oxidation cross-link Methodol.* **2014**.

(81) Kolb, H. C.; Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **2003**, *8*, 1128–1137.

(82) Tornoe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67*, 3057–3064.

(83) Valverde, I. E.; Bauman, A.; Kluba, C. A.; Vomstein, S.; Walter, M. A.; Mindt, T. L. 1,2,3-Triazoles as amide bond mimics: triazole scan yields protease-resistant peptidomimetics for tumor targeting. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 8957–8960.

(84) Echemendía, R.; Concepción, O.; Morales, F. E.; Paixão, M. W.; Rivera, D. G. The CuI-catalyzed alkyne–azide cycloaddition as direct conjugation/cyclization method of peptides to steroids. *Tetrahedron* **2014**, *70*, 3297–3305.

(85) Hollmann, A.; Matos, P. M.; Augusto, M. T.; Castanho, M. A. R. B.; Santos, N. C. Conjugation of cholesterol to HIV-1 fusion inhibitor C34 increases peptide-membrane interactions potentiating its action. *PLoS One 8*, e60302.

(86) Ingallinella, P.; Bianchi, E.; Ladwa, N. A.; Wang, Y.-J.; Hrin, R.; Veneziano, M.; Bonelli, F.; Ketas, T. J.; Moore, J. P.; Miller, M. D.; et al. Addition of a cholesterol group to an HIV-1 peptide fusion inhibitor dramatically increases its antiviral potency. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5801–5806.

(87) Tam, J. P. Recent advances in multiple antigen peptides. *J. Immunol. Methods* **1996**, *196*, 17–32.

(88) Holmes, C. P.; Jones, D. G. Reagents for Combinatorial Organic Synthesis: Development of a New o-Nitrobenzyl Photolabile Linker for Solid Phase Synthesis. *J. Org. Chem.* **1995**, *60*, 2318–2319.

(89) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem., Int. Ed.* **2009**, *48*, 9879–9883, S9879/1–S9879/10.

(90) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling Live Cells by Copper-Catalyzed Alkyne-Azide Click Chemistry. *Bioconjugate Chem.* **2010**, *21*, 1912–1916.

(91) Akoka, S.; Barantin, L.; Trierweiler, M. Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71*, 2554–2557.

(92) Villiers, A. Comptes rendus hebdomadaires des séances de l'Académie des sciences / publiés... par MM. les secrétaires perpétuels. *C. R. Hebd. Seances Acad. Sci* **1891**, *112*, 536–538.

(93) Schardinger, F. Über thermophile Bakterien aus verschiedenen Speisen und Milch. *Zeitschrift für Untersuchung der Nahrungs- und Genußm.* **1903**, *6*, 865–880.

(94) Freundenberg, K.; Cramer, F.; Plieninger, H. Inclusion Compounds of Physiologically Active Organic Compounds. 895,769, 1953.

(95) Horikoshi, K. Production and industrial application of b-cyclodextrin. *Process. Biochem.* **1979**, *14*, 26–30.

(96) Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227*, 680–685.

(97) Li, J.; Loh, X. J. Cyclodextrin-based supramolecular architectures: syntheses, structures, and applications for drug and gene delivery. *Adv. Drug Deliv. Rev.* **2008**, *60*, 1000–1017.

(98) Banerjee, S. S.; Chen, D.-H. Cyclodextrin-conjugated nanocarrier for magnetically guided delivery of hydrophobic drugs. *J. Nanoparticle Res.* **2008**, *11*, 2071–2078.

(99) Girek, T.; Goszczyński, T.; Girek, B.; Ciesielski, W.; Boratyński, J.; Rychter, P. β-Cyclodextrin/protein conjugates as a innovative drug systems: synthesis and MS investigation. *J. Incl. Phenom. Macrocycl. Chem.* **2012**, *75*, 293–296.

(100) Kamada, M.; Hirayama, F.; Udo, K.; Yano, H.; Arima, H.; Uekama, K. Cyclodextrin conjugate-based controlled release system: repeated- and prolonged-releases of ketoprofen after oral administration in rats. *J. Control. Release* **2002**, *82*, 407–416.

(101) Udo, K.; Hokonohara, K.; Motoyama, K.; Arima, H.; Hirayama, F.; Uekama, K. 5-Fluorouracil acetic acid/beta-cyclodextrin conjugates: drug release behavior in enzymatic and rat cecal media. *Int. J. Pharm.* **2010**, *388*, 95–100.

(102) Schaschke, N.; Assfalg-Machleidt, I.; Machleidt, W.; Laßleben, T.; Sommerhoff, C. P.; Moroder, L. β-Cyclodextrin/epoxysuccinyl peptide conjugates: a new drug targeting system for tumor cells. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 677–680.

(103) Decroocq, C.; Joosten, A.; Sergent, R.; Mena Barragán, T.; Ortiz Mellet, C.; Compain, P. The multivalent effect in glycosidase inhibition: probing the influence of valency, peripheral ligand structure, and topology with cyclodextrin-based iminosugar click clusters. *Chembiochem* **2013**, *14*, 2038–2049.

(104) Tsutsumi, H.; Ikeda, H.; Mihara, H.; Ueno, A. Enantioselective ester hydrolysis catalyzed by β-cyclodextrin conjugated with β-hairpin peptides. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 723–726.

(105) Hossain, M. A.; Hamasaki, K.; Takahashi, K.; Mihara, H.; Ueno, A. Guest-Induced Diminishment in Fluorescence Quenching and Molecule Sensing Ability of A Novel Cyclodextrin−Peptide Conjugate. *J. Am. Chem. Soc.* **2001**, *123*, 7435–7436.

(106) Tsutsumi, H.; Hamasaki, K.; Mihara, H.; Ueno, A. Rate enhancement and enantioselectivity in ester hydrolysis catalysed by cyclodextrin–peptide hybrids. *J. Chem. Soc. Perkin Trans. 2* **2000**, 1813–1818.

(107) Rasheed, A. Cyclodextrins as Drug Carrier Molecule: A Review. *Sci. Pharm.* **2008**, *76*, 567–598.

(108) Brewster, M. E.; Loftsson, T. Cyclodextrins as pharmaceutical solubilizers. *Adv. Drug Deliv. Rev.* **2007**, *59*, 645–666.

(109) Kumprecht, L.; Budesínský, M.; Vondrásek, J.; Vymetal, J.; Cerný, J.; Císarová, I.; Brynda, J.; Herzig, V.; Koutník, P.; Závada, J.; et al. Rigid duplex alpha-cyclodextrin reversibly connected with disulfide bonds. Synthesis and inclusion complexes. *J. Org. Chem.* **2009**, *74*, 1082–1092.

(110) Choi, S.-K. *Synthetic Multivalent Molecules: Concepts and Biomedical Applications*; John Wiley & Sons Ltd., 2005.

(111) Fulton, D. A.; Stoddart, J. F. Neoglycoconjugates Based on Cyclodextrins and Calixarenes. *Bioconjug. Chem.* **2001**, *12*, 655–672.

(112) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004–2021.

(113) Ortega-Muñoz, M.; Morales-Sanfrutos, J.; Perez-Balderas, F.; Hernandez-Mateo, F.; Giron-Gonzalez, M. D.; Sevillano-Tripero, N.; Salto-Gonzalez, R.; Santoyo-Gonzalez, F. Click multivalent neoglycoconjugates as synthetic activators in cell adhesion and stimulation of monocyte/machrophage cell lines. *Org. Biomol. Chem.* **2007**, *5*, 2291–2301.

(114) Liu, X.-M.; Lee, H.-T.; Reinhardt, R. A.; Marky, L. A.; Wang, D. Novel biomineral-binding cyclodextrins for controlled drug delivery in the oral cavity. *J. Control. Release* **2007**, *122*, 54–62.

(115) Hoogenboom, R.; Moore, B. C.; Schubert, U. S. Synthesis of star-shaped poly(epsilon-caprolactone) via'click' chemistry and "supramolecular click" chemistry. *Chem. Commun. (Camb).* **2006**, 4010–4012.

(116) Lin, Y.-C.; Wang, P.-I.; Kuo, S.-W. Water-soluble, stable helical polypeptide-grafted cyclodextrin bioconjugates: synthesis, secondary and self-assembly structures, and inclusion complex with guest compounds. *Soft Matter* **2012**, *8*, 9676.

(117) Jiménez-Balsa, A.; Pazos, E.; Martínez-Albardonedo, B.; Mascareñas, J. L.; Vázquez, M. E. Temporary electrostatic impairment of DNA recognition: light-driven DNA binding of peptide dimers. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8825–8829.

(118) Mosquera, J.; Jimanez-Balsa, A.; Dodero, V. I.; Vazquez, M. E.; Mascarenas, J. L. Stimuli-responsive selection of target DNA sequences by synthetic bZIP peptides. *Nat. Commun.* **2013**, *4*, ncomms2825, 8 pp.

(119) Hellman, L. M.; Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007**, *2*, 1849–1861.

(120) *Cell Biology (Page 330)*; Third.; Elsevier: Oxford (UK), 2006; p. 584.

(121) Hollenbeck, J. J.; Oakley, M. G. GCN4 Binds with High Affinity to DNA Sequences Containing a Single Consensus Half-Site †. *Biochemistry* **2000**, *39*, 6380–6389.

(122) Bird, G. H.; Lajmi, A. R.; Shin, J. A. Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy. *Biopolymers* **2002**, *65*, 10–20.

(123) Chan, I.-S.; Fedorova, A. V; Shin, J. a. The GCN4 bZIP targets noncognate gene regulatory sequences: quantitative investigation of binding at full and half sites. *Biochemistry* **2007**, *46*, 1663–1671.

(124) Faugeras, P.-A.; Boëns, B.; Elchinger, P.-H.; Brouillette, F.; Montplaisir, D.; Zerrouki, R.; Lucas, R. When Cyclodextrins Meet Click Chemistry. *European J. Org. Chem.* **2012**, *2012*, 4087–4105.

(125) Li, J.; Zhou, Z.; Chen, G.; Li, Q. Hierarchical Assembly of Amphiphilic POSS-Cyclodextrin Molecules and Azobenzene End-Capped Polymers. *Macromolecules* **2014**, *47*, 5739–5748.

(126) Zhang, Q.; Su, L.; Collins, J.; Chen, G.; Wallis, R.; Mitchell, D. A.; Haddleton, D. M.; Becer, C. R. Dendritic Cell Lectin-Targeting Sentinel-like Unimolecular Glycoconjugates To Release an Anti-HIV Drug. *J. Am. Chem. Soc.* **2014**, *136*, 4325–4332.

(127) Talanian, R. V; McKnight, C. J.; Kim, P. S. Sequence-specific DNA binding by a short peptide dimer. *Sci. (Washington, D. C., 1883-)* **1990**, *249*, 769–771.

(128) Bechara, C.; Sagan, S. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* **2013**, *587*, 1693–1702.

(129) Copolovici, D. M.; Langel, K.; Eriste, E.; Langel, Ü. Cell-penetrating peptides: design, synthesis, and applications. *ACS Nano* **2014**, *8*, 1972–1994.

(130) Verdine, G. L.; Hilinski, G. J. Stapled peptides for intracellular drug targets. *Methods Enzymol.* **2012**, *503*, 3–33.

(131) New, R. R. C.; Travers, G. Bile acids and biguanides as protease inhibitors for preserving the integrity of peptides in the gut, April 08, 2010.

(132) Zhao, L.; Ding, J.; He, P.; Xiao, C.; Tang, Z.; Zhuang, X.; Chen, X. An efficient pH sensitive oral insulin delivery system enhanced by deoxycholic acid. *J. Control. Release* **2011**, *152 Suppl* , e184–e186.

(133) Futaki, S.; Suzuki, T.; Ohashi, W.; Yagami, T.; Tanaka, S.; Ueda, K.; Sugiura, Y. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **2001**, *276*, 5836–5840.

(134) Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983**, *65*, 55–63.

(135) Pazos, E.; Mosquera, J.; Vázquez, M. E.; Mascareñas, J. L. DNA recognition by synthetic constructs. *Chembiochem* **2011**, *12*, 1958–1973.

(136) Johnson, I. D. Probes for Organelles. In *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*; Life Technologies Corporation, 2010.

(137) Johnson, I. D. Fluorescent Tracers of Cell Morphology and Fluid Flow. In *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*; Life Technologies Corporation, 2010.

(138) Johnson, I. D. Fluorophores and Their Amine-Reactive Derivatives. In *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies*; Life Technologies Corporation, 2010.

(139) Macey, M. G. Principles of Flow Cytometry. In *Flow cytometry: principals and application*; Humana Press: Totowa, NJ, 2007; pp. 1–15.

(140) Frankel, A. D.; Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **1988**, *55*, 1189–1193.

(141) Langel, U. The Import Mechanism of Cationic Cell-Penetrating Peptides and Its Implications for Delivery of Peptide Inhibitors of Signal Transduction. In *Handbook of Cell-Penetrating peptides*;

Fotin-Mleczek, M.; Voss, S.; Brock, R., Eds.; CRC press Taylor and Francis group: Boca Raton, 2010; pp. 161–182.

(142) Mayor, S.; Pagano, R. E. Pathways of clathrin-independent endocytosis. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 603–612.

(143) Copolovici, D. M.; Langel, K.; Eriste, E.; Langel, Ü. Cell-penetrating peptides: design, synthesis, and applications. *ACS Nano* **2014**, *8*, 1972–1994.

(144) Wadia, J. S.; Stan, R. V; Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **2004**, *10*, 310–315.

(145) Langel, U. Tat-Mediated Peptide/Protein Transduction In Vivo. In *Handbook of CellPenetrating peptides*; Shi, W.; Dowdy, S. F., Eds.; CRC press Taylor and Francis group: Boca Raton, 2007; pp. 201–217.

(146) Mukherjee, S.; Ghosh, R. N.; Maxfield, F. R. Endocytosis. *Physiol. Rev.* **1997**, *77*, 759–803.

(147) Conner, S. D.; Schmid, S. L. Regulated portals of entry into the cell. *Nature* **2003**, *422*, 37–44.

(148) Luttrell, L. M.; Daaka, Y.; Lefkowitz, R. J. Regulation of tyrosine kinase cascades by G-protein-coupled receptors. *Curr. Opin. Cell Biol.* **1999**, *11*, 177–183.

(149) Ter-Avetisyan, G.; Tünnemann, G.; Nowak, D.; Nitschke, M.; Herrmann, A.; Drab, M.; Cardoso, M. C. Cell entry of arginine-rich peptides is independent of endocytosis. *J. Biol. Chem.* **2009**, *284*, 3370–3378.

(150) Thomsen, P.; Roepstorff, K.; Stahlhut, M.; van Deurs, B. Caveolae are highly immobile plasma membrane microdomains, which are not involved in constitutive endocytic trafficking. *Mol. Biol. Cell* **2002**, *13*, 238–250.

(151) Hansen, C. G.; Nichols, B. J. Molecular mechanisms of clathrin-independent endocytosis. *J. Cell Sci.* **2009**, *122*, 1713–1721.

(152) Vivès, E.; Richard, J.-P.; Rispal, C.; Lebleu, B. TAT peptide internalization: seeking the mechanism of entry. *Curr. Protein Pept. Sci.* **2003**, *4*, 125–132.

(153) Maiolo, J. R.; Ferrer, M.; Ottinger, E. A. Effects of cargo molecules on the cellular uptake of arginine-rich cell-penetrating peptides. *Biochim. Biophys. Acta* **2005**, *1712*, 161–172.

(154) Keller, A.-A.; Mussbach, F.; Breitling, R.; Hemmerich, P.; Schaefer, B.; Lorkowski, S.; Reissmann, S. Relationships between Cargo, Cell Penetrating Peptides and Cell Type for Uptake of Non-Covalent Complexes into Live Cells. *Pharmaceuticals (Basel).* **2013**, *6*, 184–203.

(155) Lättig-Tünnemann, G.; Prinz, M.; Hoffmann, D.; Behlke, J.; Palm-Apergi, C.; Morano, I.; Herce, H. D.; Cardoso, M. C. Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat. Commun.* **2011**, *2*, 453.

(156) Lundberg, P.; Langel, U. A brief introduction to cell-penetrating peptides. *J. Mol. Recognit.* **2003**, *16*, 227–233.

(157) Beevers, A. J.; Dixon, A. M. Helical membrane peptides to modulate cell function. *Chem. Soc. Rev.* **2010**, *39*, 2146–2157.

(158) Milletti, F. Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discov. Today* **2012**, *17*, 850–860.

(159) Erazo-Oliveras, A.; Muthukrishnan, N.; Baker, R.; Wang, T.-Y.; Pellois, J.-P. Improving the endosomal escape of cell-penetrating peptides and their cargos: strategies and challenges. *Pharmaceuticals (Basel).* **2012**, *5*, 1177–1209.

(160) Gruenberg, J.; van der Goot, F. G. Mechanisms of pathogen entry through the endosomal compartments. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 495–504.

(161) Miller, D. K.; Griffiths, E.; Lenard, J.; Firestone, R. A. Cell killing by lysosomotropic detergents. *J. Cell Biol.* **1983**, *97*, 1841–1851.

(162) Varkouhi, A. K.; Scholte, M.; Storm, G.; Haisma, H. J. Endosomal escape pathways for delivery of biologicals. *J. Control. Release* **2011**, *151*, 220–228.

(163) Abes, S.; Turner, J. J.; Ivanova, G. D.; Owen, D.; Williams, D.; Arzumanov, A.; Clair, P.; Gait, M. J.; Lebleu, B. Efficient splicing correction by PNA conjugation to an R6-Penetratin delivery peptide. *Nucleic Acids Res.* **2007**, *35*, 7396–7396.

(164) Parente, R. A.; Nir, S.; Szoka, F. C. Mechanism of leakage of phospholipid vesicle contents induced by the peptide GALA. *Biochemistry* **1990**, *29*, 8720–8728.

(165) Lättig-Tünnemann, G.; Prinz, M.; Hoffmann, D.; Behlke, J.; Palm-Apergi, C.; Morano, I.; Herce, H. D.; Cardoso, M. C. Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat. Commun.* **2011**, *2*, 453.

(166) Tünnemann, G.; Ter-Avetisyan, G.; Martin, R. M.; Stöckl, M.; Herrmann, A.; Cardoso, M. C. Live-cell analysis of cell penetration ability and toxicity of oligo-arginines. *J. Pept. Sci.* **2008**, *14*, 469–476.

(167) Darnell Jr., J. E. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 740–749.

(168) Nair, S. K.; Burley, S. K. X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. *Cell (Cambridge, MA, U. S.)* **2003**, *112*, 193–205.

(169) Canne, L. E.; Ferre- D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc-Max. *J. Am. Chem. Soc.* **1995**, *117*, 2998–3007.

(170) Palmer, C. R.; Sloan, L. S.; Adrian Jr., J. C.; Cuenoud, B.; Paolella, D. N.; Schepartz, A. DNA Bending and Binding by Metallo-Zipper Models of bZIP Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 8899–8907.

(171) Morii, T.; Simomura, M.; Morimoto, S. Sequence-Specific DNA Binding by a Geometrically Constrained Peptide Dimer Sequence-specific DNA binding of these four dimeric peptides Highly Selective and Operationally Simple Synthesis of Enantiomerically Pure @ -Amino Esters via Double Stereodifferent. **1993**, *32*, 1150–1151.

(172) Verzele, D.; Madder, A. Synthetic Progress in cMyc-Max Oncoprotein Miniaturization: Semi-Online Monitoring Gives Solid-Phase Access to Hydrophobic b(-HLH-)ZIP Peptidosteroid Tweezers (Eur. J. Org. Chem. 4/2013). *Eur. J. Org. Chem.* **2013**, *2013*, n/a.

(173) Sievänen, E. Exploitation of bile acid transport systems in prodrug design. *Molecules* **2007**, *12*, 1859–1889.

(174) Neises, B.; Steglich, W. Simple Method for the Esterification of Carboxylic Acids. *Angew. Chemie Int. Ed. English* **1978**, *17*, 522–524.

(175) Höfle, G.; Steglich, W.; Vorbrüggen, H. 4-Dialkylaminopyridines as Highly Active Acylation Catalysts. [New synthetic method (25)]. *Angew. Chemie Int. Ed. English* **1978**, *17*, 569–583.

(176) HÖFLE, G.; STEGLICH, W. 4-Dialkylaminopyridines as Acylation Catalysts; III 1 . Acylation of Sterically Hindered Alcohols. *Synthesis (Stuttg).* **1972**, *1972*, 619–621.

(177) Fischer, E. Untersuchungen über Aminosäuren, Polypeptide und Proteïne. *Berichte der Dtsch. Chem. Gesellschaft* **1906**, *39*, 530–610.

(178) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

(179) Merrifield, B. The chemical synthesis of proteins. *Protein Sci.* **1996**, *5*, 1947–1951.

(180) Williams, M. J.; Muir, T. W.; Ginsberg, M. H.; Kent, S. B. H. Total Chemical Synthesis of a Folded .beta.-Sandwich Protein Domain: An Analog of the Tenth Fibronectin Type 3 Module. *J. Am. Chem. Soc.* **1994**, *116*, 10797–10798.

(181) Figaroli, S.; Madder, A. Design and automated generation of artificial estrogen receptor as potential endocrine disruptor chemical binders. *Tetrahedron* **2010**, *66*, 6912–6918.

(182) Bodé, C. a; Bechet, T.; Prodhomme, E.; Gheysen, K.; Gregoir, P.; Martins, J. C.; Muller, C. P.; Madder, A. Towards the conformational mimicry of the measles virus HNE loop: design, synthesis and biological evaluation of a cyclic bile acid-peptide conjugate. *Org. Biomol. Chem.* **2009**, *7*, 3391–3399.

(183) Hamuro, Y.; Scialdone, M. A.; DeGrado, W. F. Resin-to-Resin Acyl- and Aminoacyl-Transfer Reactions Using Oxime Supports. *J. Am. Chem. Soc.* **1999**, *121*, 1636–1644.

(184) Thompson, K. A.; Hall, D. G. Resin-to-resin Petasis borono-Mannich reaction between dialkylamino resins and supported boronic acids. *Chem. Commun.* **2000**, 2379–2380.

(185) Gravel, M.; Bérubé, C. D.; Hall, D. G. Resin-to-Resin Suzuki Coupling of Solid Supported Arylboronic Acids. *J. Comb. Chem.* **2000**, *2*, 228–231.

(186) Martos, V.; Castreño, P.; Valero, J.; de Mendoza, J. Binding to protein surfaces by supramolecular multivalent scaffolds. *Curr. Opin. Chem. Biol.* **2008**, *12*, 698–706.

(187) Uhlenheuer, D. A.; Petkau, K.; Brunsveld, L. *Combining supramolecular chemistry with biology.*; The Royal Society of Chemistry, 2010; Vol. 39, pp. 2817–2826.

(188) Battle, C. H.; Jayawickramarajah, J. Supramolecular Approaches for Inhibition of Protein–Protein and Protein–DNA Interactions. *Supramolecular Chemistry: From Molecules to Nanomaterials*. John Wiley & Sons, Ltd: Chichester, UK March 15, 2012, pp. 1885–1908.

(189) Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. The safety catch principle in solid phase peptide synthesis. *J. Chem. Soc. D Chem. Commun.* **1971**, 636.

(190) Backes, B. J.; Ellman, J. a. An Alkanesulfonamide "Safety-Catch" Linker for Solid-Phase Synthesis. *J. Org. Chem.* **1999**, *64*, 2322–2330.

(191) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. Activation Method to Prepare a Highly Reactive Acylsulfonamide "Safety-Catch" Linker for Solid-Phase Synthesis 1. *J. Am. Chem. Soc.* **1996**, *118*, 3055–3056.

(192) Quarrell, R.; Claridge, T. D. W.; Weaver, G. W.; Lowe, G. Structure and properties of TentaGel resin beads: Implications for combinatorial library chemistry. *Mol. Divers.* **1996**, *1*, 223–232.

(193) Mezzato, S.; Schaffrath, M.; Unverzagt, C. An orthogonal double-linker resin facilitates the efficient solid-phase synthesis of complex-type N-glycopeptide thioesters suitable for native chemical ligation. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1650–1654.

(194) Khattab, S. N.; Subirós-Funosas, R.; El-Faham, A.; Albericio, F. Screening of N-alkyl-cyanoacetamido oximes as substitutes for N-hydroxysuccinimide. *ChemistryOpen* **2012**, *1*, 147–152.

(195) Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R. The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4052–4055.

(196) Thompson, J. E.; Kutateladze, T. G.; Schuster, M. C.; Venegas, F. D.; Messmore, J. M.; Raines, R. T. Limits to Catalysis by Ribonuclease A. *Bioorg. Chem.* **1995**, *23*, 471–481.

(197) Cleland, W. W.; Hengge, A. C. Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev.* **2006**, *106*, 3252–3278.

(198) Roberts, R. J. How restriction enzymes became the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5905–5908.

(199) Radzicka, A.; Wolfenden, R. A proficient enzyme. *Science (80-.).* **1995**, *267*, 90–93.

(200) Sugiura, Y.; Shiraki, T.; Konishi, M.; Oki, T. DNA intercalation and cleavage of an antitumor antibiotic dynemicin that contains anthracycline and enediyne cores. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3831–3835.

(201) Long, B. H.; Golik, J.; Forenza, S.; Ward, B.; Rehfuss, R.; Dabrowiak, J. C.; Catino, J. J.; Musial, S. T.; Brookshire, K. W.; Doyle, T. W. Esperamicins, a class of potent antitumor antibiotics: mechanism of action. *Proc. Natl. Acad. Sci.* **1989**, *86*, 2–6.

(202) Chen, J.; Stubbe, J. Bleomycins: towards better therapeutics. *Nat. Rev. Cancer* **2005**, *5*, 102–112.

(203) Samejima, K.; Earnshaw, W. C. Trashing the genome: the role of nucleases during apoptosis. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 677–688.

(204) Fukui, K. The acridine ring selectively intercalated into a DNA helix at various types of abasic sites: double strand formation and photophysical properties. *Nucleic Acids Res.* **1996**, *24*, 3962–3967.

(205) Thatcher, R. J. G.; Kluger, R. *Advances in Physical Organic Chemistry Volume 25*; Advances in Physical Organic Chemistry; Elsevier, 1989; Vol. 25, pp. 99–265.

(206) Oivanen, M.; Kuusela, S.; Lönnberg, H. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brønsted Acids and Bases. *Chem. Rev.* **1998**, *98*, 961–990.

(207) Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **1980**, *49*, 877–919.

(208) Westheimer, F. H. Pseudo-rotation in the hydrolysis of phosphate esters. *Acc. Chem. Res.* **1968**, *1*, 70–78.

(209) Wolfenden, R. Degrees of difficulty of water-consuming reactions in the absence of enzymes. *Chem. Rev.* **2006**, *106*, 3379–3396.

(210) Cleland, W. W. Secondary 18O isotope effects as a tool for studying reactions of phosphate mono-, di-, and triesters. *FASEB J* **1990**, *4*, 2899–2905.

(211) Williams, N. H.; Takasaki, B.; Wall, M.; Chin, J. Structure and Nuclease Activity of Simple Dinuclear Metal Complexes: Quantitative Dissection of the Role of Metal Ions. *Acc. Chem. Res.* **1999**, *32*, 485–493.

(212) Sträter, N.; Lipscomb, W. N.; Klabunde, T.; Krebs, B. Two-Metal Ion Catalysis in Enzymatic Acyl- and Phosphoryl-Transfer Reactions. *Angew. Chemie Int. Ed. English* **1996**, *35*, 2024–2055.

(213) Wilcox, D. E. Binuclear Metallohydrolases. *Chem. Rev.* **1996**, *96*, 2435–2458.

(214) Champoux, J. J. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* **2001**, *70*, 369–413.

(215) Komiyama, M.; Shiiba, T.; Kodama, T.; Takeda, N.; Sumaoka, J.; Yashiro, M. DNA Hydrolysis by Cerium(IV) Does Not Involve either Molecular Oxygen or Hydrogen Peroxid. *Chem. Lett.* **1994**, 1025–1028.

(216) Takasaki, B. K.; Chin, J. Cleavage of the Phosphate Diester Backbone of DNA with Cerium(III) and Molecular Oxygen. *J. Am. Chem. Soc.* **1994**, *116*, 1121–1122.

(217) Komiyama, M.; Takeda, N.; Takahashi, Y.; Uchida, H.; Shiiba, T.; Kodama, T.; Yashiro, M. Efficient and oxygen-independent hydrolysis of single-stranded DNA by cerium(IV) ion. *J. Chem. Soc. Perkin Trans. 2* **1995**, 269–274.

(218) Sumaoka, J.; Azuma, Y.; Komiyama, M. Enzymatic Manipulation of the Fragments Obtained by Cerium(IV)-Induced DNA Scission: Characterization of Hydrolytic Termini. *Chem. - A Eur. J.* **1998**, *4*, 205–209.

(219) Dixon, N. E.; Geue, R. J.; Lambert, J. N.; Moghaddas, S.; Pearce, D. A.; Sargeson, A. M. DNA hydrolysis by stable metal complexes. *Chem. Commun.* **1996**, 1287–1288.

(220) Hegg, E. L.; Burstyn, J. N. Copper(II) Macrocycles Cleave Single-Stranded and Double-Stranded DNA under Both Aerobic and Anaerobic Conditions. *Inorg. Chem.* **1996**, *35*, 7474–7481.

(221) Branum, M. E.; Tipton, A. K.; Zhu, S.; Que, L. Double-Strand Hydrolysis of Plasmid DNA by Dicerium Complexes at 37 °C. *J. Am. Chem. Soc.* **2001**, *123*, 1898–1904.

(222) Schnaith, L. M.; Hanson, R. S.; Que, L. Double-stranded cleavage of pBR322 by a diiron complex via a "hydrolytic" mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 569–573.

(223) Takasaki, B. K.; Chin, J. Synergistic effect between lanthanum(III) and hydrogen peroxide in phosphate diester cleavage. *J. Am. Chem. Soc.* **1993**, *115*, 9337–9338.

(224) Liu, C.; Yu, S.; Li, D.; Liao, Z.; Sun, X.; Xu, H. DNA Hydrolytic Cleavage by the Diiron(III) Complex Fe 2 (DTPB)(μ-O)(μ-Ac)Cl(BF 4) 2 : Comparison with Other Binuclear Transition Metal Complexes. *Inorg. Chem.* **2002**, *41*, 913–922.

(225) Zhu, L.; dos Santos, O.; Koo, C. W.; Rybstein, M.; Pape, L.; Canary, J. W. Geometry-dependent phosphodiester hydrolysis catalyzed by binuclear copper complexes. *Inorg. Chem.* **2003**, *42*, 7912–7920.

(226) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B. Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.* **2004**, *126*, 10339–10349.

(227) Schmitz, K.; Schepers, U. Polyamides as artificial transcription factors: novel tools for molecular medicine? *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 2472–2475.

(228) Li, C.; Qiao, R.-Z.; Wang, Y.-Q.; Zhao, Y.-F.; Zeng, R. Synthesis and biological evaluation of the Zn (II)-IDB complexes appended with oligopolyamide as potent artificial nuclease. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5766–5770.

(229) Kameshima, W.; Ishizuka, T.; Minoshima, M.; Yamamoto, M.; Sugiyama, H.; Xu, Y.; Komiyama, M. Conjugation of peptide nucleic acid with a pyrrole/imidazole polyamide to specifically recognize and cleave DNA. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 13681–13684.

(230) Kovacic, R. T.; Welch, J. T.; Franklin, S. J. Sequence-selective DNA cleavage by a chimeric metallopeptide. *J. Am. Chem. Soc.* **2003**, *125*, 6656–6662.

(231) Nomura, A.; Sugiura, Y. Sequence-selective and hydrolytic cleavage of DNA by zinc finger mutants. *J. Am. Chem. Soc.* **2004**, *126*, 15374–15375.

(232) Komiyama, M. Sequence-Specific and Hydrolytic Scission of DNA and RNA by Lanthanide Complex-OligoDNA Hybrids. *J. Biochem.* **1995**, *118*, 665–670.

(233) Zelder, F. H.; Mokhir, A. A.; Krämer, R. Sequence selective hydrolysis of linear DNA using conjugates of Zr(IV) complexes and peptide nucleic acids. *Inorg. Chem.* **2003**, *42*, 8618–8620.

(234) Kitamura, Y.; Komiyama, M. Preferential hydrolysis of gap and bulge sites in DNA by Ce(IV)/EDTA complex. *Nucleic Acids Res.* **2002**, *30*, e102.

(235) Chen, W.; Kitamura, Y.; Zhou, J.-M.; Sumaoka, J.; Komiyama, M. Site-selective DNA hydrolysis by combining Ce(IV)/EDTA with monophosphate-bearing oligonucleotides and enzymatic ligation of the scission fragments. *J. Am. Chem. Soc.* **2004**, *126*, 10285–10291.

(236) Li, C.; Du, C.; Tian, H.; Jiang, C.; Du, M.; Liu, Y.; Qiao, R.-Z.; Jia, Y.-X.; Zhao, Y.-F. Artificial transcription factors which mediate double-strand DNA cleavage. *Chemistry* **2010**, *16*, 12935–12940.

(237) Kameshima, W.; Ishizuka, T.; Minoshima, M.; Yamamoto, M.; Sugiyama, H.; Xu, Y.; Komiyama, M. Conjugation of peptide nucleic acid with a pyrrole/imidazole polyamide to specifically recognize and cleave DNA. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 13681–13684.

(238) Feng, G.; Natale, D.; Prabaharan, R.; Mareque-Rivas, J. C.; Williams, N. H. Efficient phosphodiester binding and cleavage by a ZnII complex combining hydrogen-bonding interactions and double Lewis acid activation. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7056–7059.

(239) Feng, G.; Mareque-Rivas, J. C.; Torres Martín de Rosales, R.; Williams, N. H. A highly reactive mononuclear Zn(II) complex for phosphodiester cleavage. *J. Am. Chem. Soc.* **2005**, *127*, 13470–13471.

(240) Forconi, M.; Williams, N. H. Mimicking metallophosphatases: revealing a role for an OH group with no libido. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 849–852.

(241) Verzele, D.; Figaroli, S.; Madder, A. Shortcut access to peptidosteroid conjugates: building blocks for solid-phase bile acid scaffold decoration by convergent ligation. *Molecules* **2011**, *16*, 10168–10186.

(242) Figaroli, S.; Madder, A. Development of a synthetic receptor for trace determination of Endocrine Disruptor Chemicals, Ghent University, 2010.

(243) Thomas, P.; Smith, J. Binding of xenobiotics to the estrogen receptor of spotted seatrout: A screening assay for potential estrogenic effects. *Mar. Environ. Res.* **1993**, *35*, 147–151.

(244) Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753–758.

(245) Davis, A.; Wareham, R. Carbohydrate Recognition through Noncovalent Interactions: A Challenge for Biomimetic and Supramolecular Chemistry. *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 2978–2996.

(246) Tamminen, J.; Kolehmainen, E. Bile Acids as Building Blocks of Supramolecular Hosts. *Molecules* **2001**, *6*, 21–46.

(247) Díaz Velázquez, H.; Ruiz García, Y.; Vandichel, M.; Madder, A.; Verpoort, F. Water-soluble NHC-Cu catalysts: applications in click chemistry, bioconjugation and mechanistic analysis. *Org. Biomol. Chem.* **2014**, *12*, 9350–9356.

(248) verpoort, francis; Díaz Velázquez, H.; Ruiz García, Y.; Vandichel, M.; Madder, A. Water-soluble NHC-Cu Catalysts: Applications in Click Chemistry, Bioconjugation and Mechanistic Analysis. *Org. Biomol. Chem.* **2014**.

(249) Meldal, M.; Tornøe, C. W. Cu-catalyzed azide-alkyne cycloaddition. *Chem. Rev.* **2008**, *108*, 2952–3015.

(250) Kuang, G.-C.; Guha, P. M.; Brotherton, W. S.; Simmons, J. T.; Stankee, L. A.; Nguyen, B. T.; Clark, R. J.; Zhu, L. Experimental investigation on the mechanism of chelation-assisted, copper(II) acetate-accelerated azide-alkyne cycloaddition. *J. Am. Chem. Soc.* **2011**, *133*, 13984–14001.

(251) Adzima, B. J.; Tao, Y.; Kloxin, C. J.; DeForest, C. A.; Anseth, K. S.; Bowman, C. N. Spatial and temporal control of the alkyne-azide cycloaddition by photoinitiated Cu(II) reduction. *Nat. Chem.* **2011**, *3*, 256–259.

(252) Díez-González, S.; Correa, A.; Cavallo, L.; Nolan, S. P. (NHC)Copper(I)-catalyzed [3+2] cycloaddition of azides and mono- or disubstituted alkynes. *Chemistry* **2006**, *12*, 7558–7564.

(253) Díez-González, S.; Nolan, S. P. [(NHC)2Cu]X complexes as efficient catalysts for azide-alkyne click chemistry at low catalyst loadings. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 8881–8884.

(254) Chan, T. R.; Fokin, V. V. Polymer-Supported Copper(I) Catalysts for the Experimentally Simplified Azide–Alkyne Cycloaddition. *QSAR Comb. Sci.* **2007**, *26*, 1274–1279.

(255) Gaulier, C.; Hospital, A.; Legeret, B.; Delmas, A. F.; Aucagne, V.; Cisnetti, F.; Gautier, A. A water soluble Cu(I)-NHC for CuAAC ligation of unprotected peptides under open air conditions. *Chem. Commun. (Camb).* **2012**, *48*, 4005–4007.

(256) Stadtman, E. R. Protein oxidation and aging. *Free Radic. Res.* **2006**, *40*, 1250–1258.

(257) Stadtman, E. R.; Berlett, B. S. Reactive oxygen-mediated protein oxidation in aging and disease. *Drug Metab. Rev.* **1998**, *30*, 225–243.

(258) Kumar, A.; Li, K.; Cai, C. Anaerobic conditions to reduce oxidation of proteins and to accelerate the copper-catalyzed "Click" reaction with a water-soluble bis(triazole) ligand. *Chem. Commun. (Camb).* **2011**, *47*, 3186–3188.

(259) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **2004**, *6*, 2853–2855.

(260) Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. Minimum length of a sequence-specific DNA binding peptide. *Biochemistry* **1992**, *31*, 6871–6875.

(261) Langel, U. Tat-Derived Cell-Penetrating Peptides: Discovery, Mechanism of Cell Uptake, and Applications to the Delivery of Oligonucleotides. In *Handbook of Cell Penetrating peptides*; S. Abes, et al., Ed.; CRC press Taylor and Francis group: Boca Raton, 2007; pp. 29–42.

(262) Beal, D. M.; Albrow, V. E.; Burslem, G.; Hitchen, L.; Fernandes, C.; Lapthorn, C.; Roberts, L. R.; Selby, M. D.; Jones, L. H. Click-enabled heterotrifunctional template for sequential bioconjugations. *Org. Biomol. Chem.* **2012**, *10*, 548–554.

(263) Grishina, A.; Stanchev, S.; Kumprecht, L.; Buděšínský, M.; Pojarová, M.; Dušek, M.; Rumlová, M.; Křížová, I.; Rulíšek, L.; Kraus, T. β-Cyclodextrin duplexes that are connected through two disulfide bonds: potent hosts for the complexation of organic molecules. *Chemistry* **2012**, *18*, 12292–12304.

(264) Volkov, S.; Kumprecht, L.; Buděšínský, M.; Dušek, M.; Kraus, T. No Title. *To be Submitt.*

(265) Van Lysebetten, D. Design and Synthesis of Stapled Peptides for Enhanced DNA Binding and Cell Penetration, Ghent University, 2014.

Annex I:

Published articles

Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c4ob01350f

Water-soluble NHC-Cu catalysts: applications in click chemistry, bioconjugation and mechanistic analysis†

Heriberto Díaz Velázquez,[a,b] Yara Ruiz García,[c] Matthias Vandichel,[d] Annemieke Madder[c] and Francis Verpoort*[a,b,e]

Received 29th June 2014,
Accepted 4th September 2014

DOI: 10.1039/c4ob01350f

www.rsc.org/obc

Copper(I)-catalyzed 1,3-dipolar cycloaddition of azides and terminal alkynes (CuAAC), better known as "click" reaction, has triggered the use of 1,2,3-triazoles in bioconjugation, drug discovery, materials science and combinatorial chemistry. Here we report a new series of water-soluble catalysts based on N-heterocyclic carbene (NHC)-Cu complexes which are additionally functionalized with a sulfonate group. The complexes show superior activity towards CuAAC reactions and display a high versatility, enabling the production of triazoles with different substitution patterns. Additionally, successful application of these complexes in bioconjugation using unprotected peptides acting as DNA binding domains was achieved for the first time. Mechanistic insight into the reaction mechanism is obtained by means of state-of-the-art first principles calculations.

Introduction

Although several catalysts have been developed for the Copper catalyzed Alkyne–Azide Cycloaddition reactions (CuAAC, better known as click reaction), in most cases they need the presence of co-catalysts such as auxiliary ligands, bases – mainly amines –, and reducing or oxidizing agents depending on the Cu source used, in order to enhance their catalytic activity, although in recent years single Cu salts have also been successful.[1] Generally high Cu loading must be applied to ensure good performance of the catalyst.[2] Several Cu(I) complexes holding N-heterocyclic carbene (NHC) ligands are reported as catalysts for the Huisgen cycloaddition reaction; however, the catalysis is carried out at elevated temperatures and in the presence of organic solvents, under two-phase systems (when both reactants are not soluble) and only the 1,4-disubstituted triazoles are generated.[3,4] Moreover, their activity in the solution phase is significantly lower than the other catalytic systems.[5] Here we report a unique Cu(I) based catalyst that enables the synthesis of triazoles with different substitution patterns, *e.g.* 1-, 4-, 1,4- or 1,4,5-substituted triazoles and at the same time, able to generate triazoles from acetylene in water, and able to produce tosyl acetamides in aqueous media (Fig. 1).

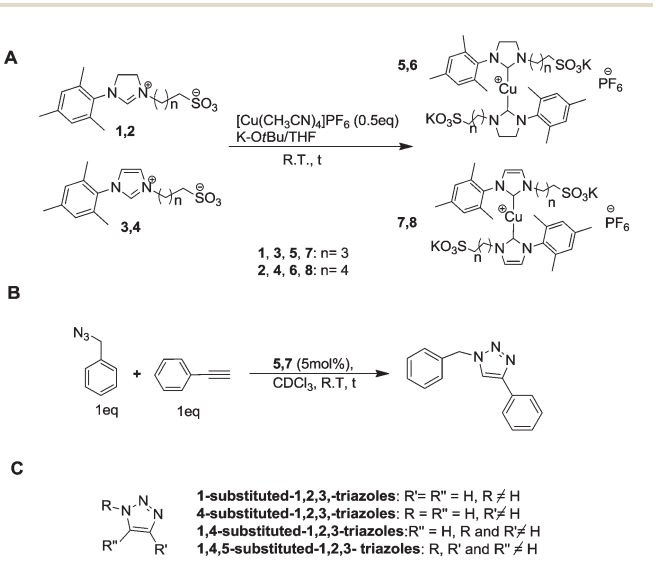


Fig. 1 (A) Synthesis of catalysts **5–8**; (B) standard click reaction monitored by ^{1}H-NMR using catalysts **5** and **7**; (C) triazoles synthesized in this work.

[a]*Laboratory of Organometallics, Catalysis and Ordered Materials, State Key Laboratory of Advanced Technology for Materials Synthesis and Processing; Center for Chemical and Material Engineering, Wuhan University of Technology; Department of Applied Chemistry, Faculty of Sciences, Wuhan University of Technology, Wuhan 430070, P.R. China. E-mail: francis.verpoort@ghent.ac.kr*
[b]*Ghent University Global Campus, Songdomunhwa, Yeonsu-Gu, Incheon, Korea*
[c]*Department of Organic Chemistry, Ghent University, Krijgslaan 281-S4, 9000 Ghent, Belgium*
[d]*Department of Molecular Modelling, Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium*
[e]*Tomsk Polytechnic University, 634050 Tomsk, Russian Federation*

† Electronic supplementary information (ESI) available. CCDC 1002517. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c4ob01350f

Until now, no such catalyst has been reported in the literature. More specifically, we report on the synthesis and application of new water-soluble $(NHC)_2Cu(I)$ complexes (**5**, **6** and **7**, **8**, Fig. 1A) in click chemistry in water, neat conditions and organic solvents or mixtures, acetylene conversion into triazoles (1-, 4-, 1,4- or 1,4,5-substituted triazoles, Fig. 1C) and acetamides along with the ease of immobilization and a detailed mechanistic analysis of the transition states involved in the catalytic process. In the last decade, bioconjugation has become a common tool for chemists and biochemists as a result of the increase in peptide and protein research and development. This covalent functionalization of these biomacromolecules under physiological conditions gives rise to new areas of research such as drug discovery, high-throughput screening and *in vivo* testing. Therefore, many ligation methodologies have been studied and optimized.[6–8] The development of click chemistry has caused an emerged interest in the functionalization of peptides with alkynes and azides, resulting in a broad variety of building blocks such as unnatural amino-acids, currently commercially available. In practice, the application of CuAAC in peptide chemistry requires extra efforts for the optimization of the reaction conditions.[8] Traces of O_2 can be reduced by Cu^I and sodium ascorbate into H_2O_2, a reactive oxygen species that can induce degradation of amino acids and cleavage of the polypeptide chain.[9–12] Thus, a catalytic system able to perform click reactions in biomacromolecules under reductant-free aqueous conditions will lead to a major step forward in bioconjugation chemistry. Furthermore, bioconjugation imposes specific requests, such as exquisite chemoselectivity, biocompatibility, and the ability to work at low temperature still enabling fast transformations. To date the CuAAC reaction in the absence of an accelerating ligand is simply too slow.[13]

In this context, we have studied the performance of our water-soluble NHC-Cu(I) catalysts for CuAAC reactions with unprotected peptides under aqueous conditions. We have chosen the DNA binding domain of the transcription factor protein GCN4, belonging to the bZIP leucine zipper family as a substrate for the reaction.[14] Indeed, modification of the GCN4 binding domain in the design of transcription factor models has been of particular interest for gene therapy.[15] Surprisingly, there are no studies so far on the application of click chemistry on the modified GCN4 binding domain for conjugation purposes. Here, we report the successful functionalization of the peptide comprising the basic region DNA binding domain with an organic azide *via* click chemistry using the sulfonated NHC-Cu(I) catalysts.

Results and discussion

Our initial experiments consisted of the suitability tests for the N-heterocyclic carbene (NHC) ligand precursors on Cu using a Cu(I) source. We envisaged that according to the literature,[4] our sulfonate-functionalized ligands should behave in a similar fashion as non-functionalized ligands. However the new compounds demonstrate an excellent behavior for click reactions in water, neat conditions and organic solvents or mixtures, for bioconjugation *via* CuAAC and for acetylene conversion into triazoles and even acetamides.

Synthesis of catalysts

For the initial step in the preparation of catalysts, NHC-ligand precursors **1–4** were synthesized (Fig. 1A). Among these, structures **1** and **2** have no precedent reports in the literature. Ligand precursors **1–4** were originally considered for a single mono-coordination at a 1 : 1 ratio using CuCl as the source of Cu^I. However, mass spectroscopy proved that the final product was actually a mixture of mono- and bis-coordinated NHC-Cu complexes. Isolation of these complexes was not feasible and therefore we were prompted to find a more convenient Cu(I) source. From literature reports, it was clear that some Cu(I) complexes with NHC ligands demonstrated a good activity of which those having PF_6^- as a counterion displaying activities more than 3 times higher than the analogous complex bearing BF_4^- as a counterion.[4] Consequently, we were encouraged to use PF_6^- as the counterion to improve the stability without affecting the catalyst activity. A synthesis protocol similar to that described in the literature was followed, using KO*t*-Bu as base and tetrahydrofuran (THF) as the solvent. Filtration and precipitation with ether resulted in a series of catalysts **5–8** depicted in Fig. 1A. $[Cu(CH_3CN)_4]PF_6$ was used as the source of Cu^I for this purpose. Characterization of complexes **5–8** by ^{13}C-Nuclear Magnetic Resonance (NMR) spectroscopy resulted in the typical values for the C1 carbene signal for saturated NHC-Cu complexes, 200.65 ppm and 200.74 ppm for **5** and **6**, respectively, while for the unsaturated NHC-Cu complexes, the values of 176.38 and 176.41 ppm were observed for **7** and **8**, respectively. Mass spectroscopic results confirmed the identity of the bis-coordinated complexes using these ligands. The newly synthesized complexes exhibited solubility in chlorinated solvents, alcohols, dimethyl sulfoxide (DMSO), dimethylformamide (DMF) and water. Their high stability to air and heat, allows storage of the compounds under an Ar atmosphere for several weeks without loss of activity.

Catalytic activity

Catalytic experiments were carried out using the standard substrates benzyl azide and phenylacetylene under neat conditions. ESI Table S1† shows the yields obtained with catalysts **5–8** proving that the catalysts bearing saturated NHC ligands are better performing than those holding unsaturated NHC ligands and also showing that catalyst **5** is highly competitive with the most active complex $[(ICy)_2Cu]PF_6$ reported by Nolan.[4] Although certain NHC-Cu(I) complexes seem to be competitive with catalyst **5**,[3,4] the special hydrophilicity of complexes **5–8** enables the extension of their scope towards different reaction conditions whereas those described in literature might hardly achieve good performance. In order to make a more detailed comparison between the newly synthesized catalysts, **5** and **7** were chosen for a reaction monitoring by ^{1}H-NMR for the standard reaction depicted in Fig. 2. That saturated NHC-Cu complexes are better performing than those holding unsaturated NHC ligands is illustrated in Fig. 2 in which **5** clearly shows a higher reactivity

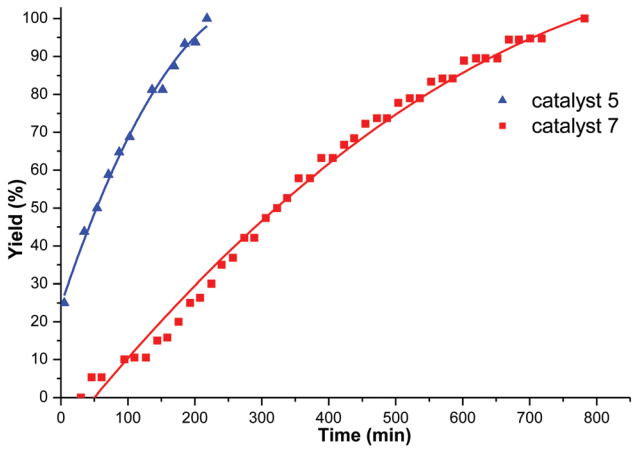


Fig. 2 Plot of the reaction yield of a standard click reaction (Fig. 1B) monitored by ^{1}H-NMR using catalysts **5** and **7**.

compared with **7** (25% conversion after the induction period of 5 min for catalyst **5** and no conversion for **7** at the same period) and a rate enhancement in the range of 3.5–4.0. The reason for this higher activity may reside in the higher σ-donating ability of the saturated NHC ligands; a similar effect causing a higher activity for NHC-Ru complexes was observed in the olefin metathesis reaction.[16–20] In view of the higher reactivity of **5** bearing sulfonated NHC ligands, these were further explored for the synthesis of triazoles with various substitution patterns.

Synthesis of triazoles

One of the major catalytic improvements made in the new complexes was the possibility to employ them for the synthesis of a variety of triazoles and the versatility of conditions. This capability is distributed from neat over solvent mixtures to *in situ* conditions. For the production of 1,4-substituted triazoles under neat conditions (ESI Table S2†) the products were pure enough for characterization by NMR spectroscopy. These varieties of triazoles were also produced *in situ*, starting from the alkyl/aryl halide and sodium azide, where the use of DMSO/water mixture was critical to achieve excellent reaction yields (ESI Table S3†). Although Ru-catalysts are more feasible for click reactions of internal alkynes the newly developed Cu-catalysts in this work were also capable of generating 1,4,5-substituted triazoles (ESI Table S4†). Due to its high-energy consumption and the use of multistep procedures,[21–23] as well as the employment of toxic solvents,[25,26] the production of 4- substituted-1,2,3-triazoles has represented a synthetic challenge. We found out the feasibility to produce 4-substituted-1,2,3-triazoles from non-activated terminal alkynes using mixtures of DMSO/AcOH and NaN_3 (ESI Table S7†), representing an improvement of the few catalytic protocols of this kind.[22–24] However, the most outstanding improvement of this work was by far the synthesis of 1-substituted triazoles from acetylene gas completely in water and in water–DMSO mixtures (Fig. 3A), alternative to those procedures involving organic solvents and higher catalyst loadings.[27,28] Finally, the reaction of tosyl azide and acetylene gas in water–DMSO mixtures yielded tosyl-acetamide (Fig. 3B), as previously reported,[29] opening the possibility of using acetylene gas as an efficient source of acetamide bonds in organic and biological chemistry.[30] All these achievements in catalytic activity were attained using a maximum of 5 mol% of the catalyst and under general aqueous conditions, which embodies an eco-friendly series of catalytic protocols.

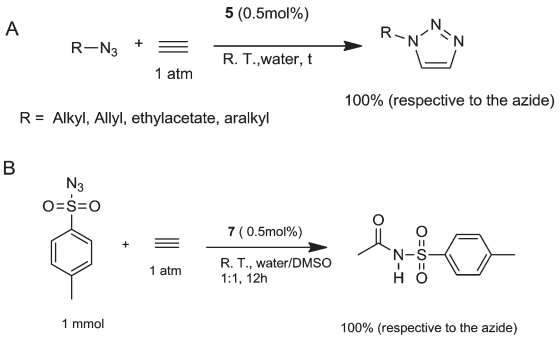


Fig. 3 (A) Use of acetylene gas in click reactions in water; (B) tosyl acetamide formation in aqueous media.

Bioconjugation experiments

To further expand and illustrate the scope and potential of the developed catalysts, bioconjugation experiments with alkyne and/or azide tagged peptides were carried out. Hereto, two peptides, one 10 mer and one 23 mer, comprising part of the N-terminal basic region of the transcription factor protein GCN4, were synthesized on 2-chlorotrityl chloride polystyrene resin. Propargyl glycine was incorporated as the first, C-terminal amino acid residue (Fig. 4) to provide for the alkyne functionality for further click reaction with benzylazide. A click reaction was carried out at the ambient temperature in a deoxygenated mixture of water and hexafluoroisopropanol to disrupt H-bonding interactions within the peptide using 1.5 equivalents of catalyst (**6** for modification of 10-mer GCN4 and **7** for 23-mer GCN4 due to its higher stability) under anaerobic conditions overnight. The resulting modified peptides were purified by High Performance Liquid Chromatography (HPLC) and further characterized by Electrospray Ionization Liquid Chromatography connected to a Mass Spectrometer (LC-MS-ESI) and MALDI-TOF. Generation of the Cu-acetylide intermediate in such an environment is highly challenging, therefore a slight excess of catalyst **6** is needed. In the current context, it is further worth noting that both sequences are rich in arginine residues, of which the guanidinium side chain can easily coordinate with copper and inhibit the reaction.

These experiments underscore the potential for application of our water-soluble $(NHC)_2Cu(I)$ catalysts **6** and **7** in the functionalization of deprotected peptides applying the CuAAC strategy (see ESI†).

Heterogeneous click catalysis by an ionically immobilized $(NHC)_2$-Cu catalyst

Further encouraged to take advantage of the ionic functionality we used this feature to ionically immobilize these com-

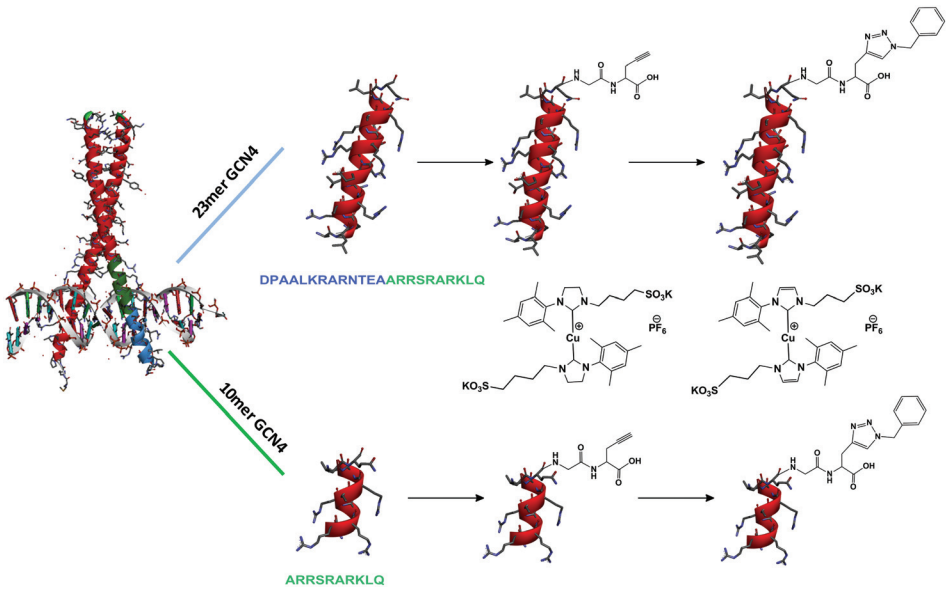


Fig. 4 The transcription factor protein GCN4, belonging to the bZIP leucine zipper family and the derived model peptide comprising the DNA binding domain. Functionalization of a peptide with an alkyne allows cycloaddition-based modification with a non-peptide azide moiety.

plexes on an anion exchange resin. For this purpose we used the crosslinked polystyrene-*co*-divinylbenzene based anion exchanger Amberlyte® IRA 402, which holds benzyl-trimethyl ammonium groups. After two washing cycles with water, an immobilized NHC-Cu(I) catalyst **6-IRA 402**, shown in Fig. 4, was obtained. Several hours were needed for the total ion exchange, which was visually detected by the colour change of the initial solution. At the end of the reaction, the solution became clear from being initially brownish, the resin embedded with the catalyst changed from its typical yellowish to brownish colour (ESI Picture S1†). The analysis of the water phase by XRF did not show any trace of Cu(I) ions after the immobilization, which confirmed the total anchoring of the Cu(NHC) catalyst on the resin.

The new heterogeneous catalyst **6-IRA 402** was applied for the three-component click reaction of benzyl azide (deriving from benzyl bromide and sodium azide) and phenylacetylene. The reaction carried out under mechanical stirring showed low conversion under standard conditions, probably due to the poor contact between the reactant and the heterogeneous catalyst. However, when applying ultrasound instead of mechanical stirring, the reaction was significantly accelerated, allowing full conversion after 5 h (see Fig. 4). This prompted us to reuse the immobilized catalyst by extracting the newly formed product and adding fresh substrates while avoiding contact of the heterogeneous catalyst with the environment. The capability of this heterogeneous catalyst to be reused allowing high yields was illustrated for a maximum of 4 catalytic cycles. After the fourth cycle, the yield decreases dramatically, possibly due to the leaching of the Cu(I) species, which was confirmed by XRF experiments of the organic phase (see ESI Graph S1†).

To our delight, the heterogeneous catalyst **6-IRA 402** was able to perform click reactions using acetylene gas as the source of alkyne. Again, the use of ultrasound was needed to achieve high conversion in a short time. This provides high versatility of these catalytic systems to be used on industrial scale (Fig. 5).

Mechanistic analysis

In order to obtain a full mechanistic understanding of the catalyst, we decided to perform state-of-the-art first principles calculations on the coupling reaction between benzylazide and phenylacetylene with two catalysts **6** and **8** (Fig. 1), bearing respectively saturated and unsaturated NHC ligands. The goal of the calculations is twofold: (1) providing a plausible reaction mechanism of CuAAC catalyzed reactions by these new water soluble Cu-NHC complexes and (2) investigating the electronic effect of the carbon–carbon backbone of the NHC-ligand on the click reaction. Full catalyst models of species **6** and **8** (Fig. 1) were modeled at the B3LYP/6-311++g(3df,2p)-D3//B3LYP/6-31+g(d) level of theory (details can be found in ESI†) and a free energy diagram was constructed along the proposed mechanistic cycle.

For the coupling between these molecules, taking place on bi-ligand complexes (Fig. 1), there are steric limitations and therefore the computations are restricted to mono-ligand versions of catalysts **6** and **8**. Under experimental conditions, one can expect that mono-ligand complexes ($LCuPF_6$) as well as bi-ligand complexes (L_2CuPF_6) are available, which is the case if there exists an equilibrium between them. The two possibilities to generate mono-ligand complexes are given in Fig. 6A. In the presence of phenylacetylene, the bi-ligand complex can also react towards a monoligand complex LCu-acetylide and $L{\cdot}HPF_6$, as was verified by Diez-Gonzalez *et al.*[3] If such a reaction occurs, LCu-acetylide can catalyze the CuAAC reaction. Remark that dependent on the counterion (in our case PF_6^-), we can expect different crystallization behavior of

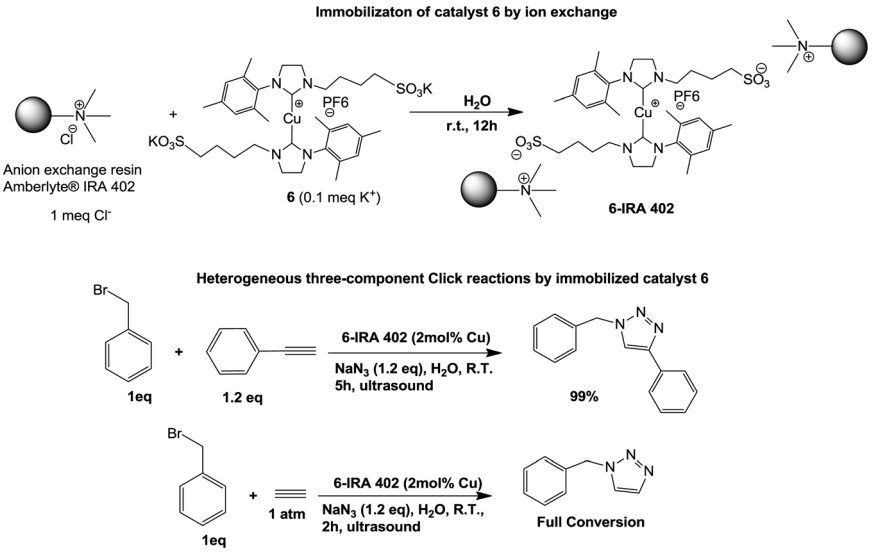


Fig. 5 Synthesis and application in heterogeneous click reactions of an ionically immobilized $(NHC)_2Cu(I)$ catalyst.

one of the ligand, and correspondingly a different amount of active LCu-acetylide complexes (Fig. 6A). However, monoligated complexes can also be generated without acetylene, through the equilibrium reaction between L_2CuPF_6 and $LCuPF_6$, in which a ligand is released (Fig. 6A). This compound can as well react with phenylacetylene forming a LH^+CuPF_6-acetylide, in which the sulfate group of the ligand has taken up the acidic proton of phenylacetylene. From the free energy differences displayed in Fig. 6A (respectively 289.9 and 163.0 kJ mol^{-1} for the saturated NHC-complex), we can assume that a reaction of the bi-ligand complex with phenylacetylene is the easiest route to a mono-ligand complex. In the following, catalytic cycles with L-Cu-acetylide as an intermediate will further be discussed, while catalytic cycles with LH^+CuPF_6-acetylide as an intermediate are described in ESI.†

The Cu(I)-acetylide complex was also observed by Nolte *et al.*, who postulated – based on experimental findings – a six-step catalytic cycle for a standard click reaction on Cu.[5] Our modeled catalytic cycles contain much similarity with their scheme; however, we have chosen to display only the two most important intermediate complexes in our scheme (Fig. 6B) neglecting the shallow intermediates. For example, we observed that the Cu-triazole product is formed immediately according to the intrinsic reaction coordinate (IRC) scan from **TS-1** (Fig. 6B (left) and 6D). The Cu-triazole can then be protonated by a co-adsorbed phenylacetylene (**TS-2**, Fig. 6B (left) and 6D), generating the 1,4-triazole product and the active Cu(I)-acetylide complex. For complex **6**, the coupling reaction has a free energy barrier of 113.0 kJ mol^{-1} and the proton transfer reaction a free energy barrier of 128.5 kJ mol^{-1} (Fig. 6B). It can however not be excluded that more stable pathways can be found computationally, in which the ligand groups interact differently with the reactants. If the reaction proceeds without the catalyst, the observed reaction barrier towards 1-benzyl-4-phenyl-1,2,3-triazole amounts to 132.9 kJ mol^{-1}, which is still higher than the highest barrier in the catalytic scheme (*cf.* Fig. 6C and B, left). Moreover, the Cu-NHC catalyst selectively generates only the 1,4-triazole product, while thermal reactions without the catalyst produce a racemic mixture (**3**) of the two possible triazoles.

Furthermore, just as in the experiments under neat conditions, the saturated NHC-complexes were found to be lower activated than the unsaturated ones. Based on the differences in the activation barrier for the rate determining step, we can have an indication for the rate acceleration of the saturated catalyst *versus* the unsaturated catalyst. Based on the coupling step (TS-1), this would be around 5.4 times faster, while based on the hydrogen transfer step (TS-2), this would be approximately 3.6 times faster at 298 K. In the first instance, this seems to be in good agreement with the experimentally observed rate acceleration of about 4 between catalysts **5** and **7** (Fig. 2). As these catalysts have only one CH_2-group less compared to catalysts **6** and **8**, the difference in the rate between catalysts **6** and **8** will be similar, and thus, is in rather good agreement with the computationally predicted acceleration. For this catalytic cycle (Fig. 6B, left), the influence of water solvation was studied in parallel (Fig. 6B, right). With implicit solvation for water, the activation barriers are remarkably higher (for the computational methodology, see ESI†). Note as well that the differences in the rate between the saturated and the unsaturated complexes become smaller with water as the solvent, leading right to acceleration factors of around 2. From an experimental point of view, it would also be interesting to determine the various rate acceleration factors between complexes **5** and **7** in many solvents, yet this clearly falls beyond the scope of this initial communication on these new water-soluble Cu-NHC-catalysts. However, for future research, this might validate the proposed catalytic cycles.

The free energy contributions between the different states of the catalytic cycle can be split into enthalpy and entropy

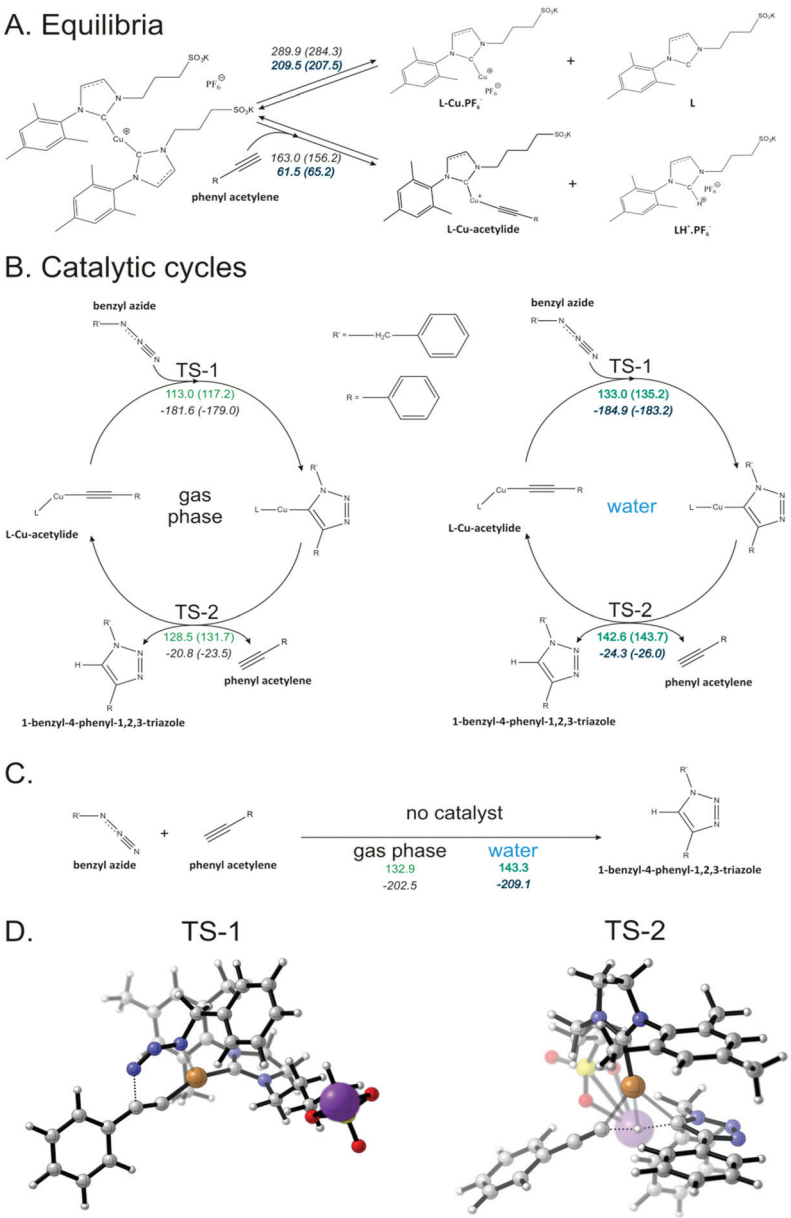


Fig. 6 (A) Equilibrium between bi-ligand and mono-ligand complexes; reaction free energies (298 K) are given in kJ mol^{-1} for the saturated (unsaturated, in parentheses) catalyst in the gas phase (top, black) and water (bottom, blue). (B) Simplified catalytic cycles in which L-Cu represents the mono-ligated Cu-NHC-complex **6** or **8**, free energy barriers (top, green) and reaction free energies (bottom, black) are given in kJ mol^{-1} at 298 K for the saturated (and for unsaturated catalyst in parentheses) catalyst. (C) Reactions between benzylazide and phenylacetylene without the catalyst. (D) Visualization of the modeled transition states TS-1 and TS-2 on catalyst model **6**.

contributions (ESI Table S7†). The click reaction is very exergonic with reaction free energy differences around −305 kJ mol^{-1}. To quantify the relative rates between catalysts **6** and **8**, one can look at the difference in activation free energy for the global barrier. Therefore, a reduced catalytic cycle is constructed with only the major states (right part, Fig. 6A), omitting the (de)protonation transition states. Right now, the largest free energy barrier of both catalytic cycles can easily be compared. The global free energy barrier from **F** amounts to 141.6 kJ mol^{-1} and 145.7 kJ mol^{-1}, for catalysts **6** and **8**, respectively. This gives a rate acceleration of 5.3 between both complexes, which is in agreement with the experimentally observed rate acceleration of 4 between catalysts **5** and **7** (Fig. 2). As these catalysts have only one CH_2 group less compared to catalysts **6** and **8**, the difference in the rate between catalysts **6** and **8** will be similar, and thus, is in very good agreement with the computationally predicted acceleration. Without a catalyst the free energy barrier would be 132.9 kJ mol^{-1}, which is larger than the barrier measured from complex **A** (90.9 and 95.8 kJ mol^{-1}, ESI Table S8†) but still lower than the barrier from complex **F** (141.6 and 145.7 kJ mol^{-1}).

Conclusion and outlook

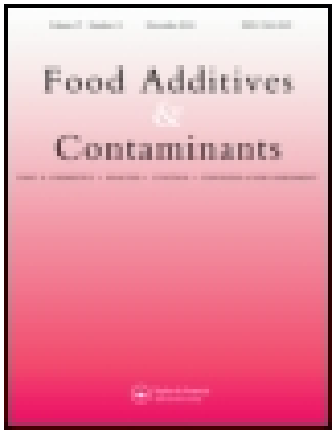
We have presented for the first time the synthesis of sulfonate functionalized bis-NHC-Cu(I) complexes and their application in click reactions for a variety of reaction conditions, illustrating the high versatility of these compounds, for the synthesis of triazoles with different substitution patterns. We have further shown the use of these water-soluble complexes in bioconjugation experiments, attaining for the first time the functionalization of an unprotected 25-aminoacid chain peptide using low catalyst loading, which otherwise would need a much higher amount of Cu(I) to achieve high efficiency due to the chelating properties of certain aminoacid residues. We could take further advantage of the ionic functionality inserted into the NHC ligand and used this feature to ionically immobilize these complexes on an anion exchange resin. Moreover, both, the homogeneous and the heterogeneous catalysts even performed click reactions using acetylene gas as the alkyne source. Exploiting tosyl azide as a substrate generated tosyl-acetamide as the sole product and thus broadens the scope of the newly developed catalysts. Mechanistic analysis of the Cu(I) catalysts bearing sulfonate functionalized NHC ligands shows that the sulfonate group allows internal deprotonation and protonation steps of the alkyne and formed reaction products, respectively. Comparison of energy barriers quantitatively confirmed that the complexes with saturated ligands are more active than the complexes with unsaturated ligands toward click reactions, which is in good agreement with the experiments.

Acknowledgements

F.V. would like to express his deep appreciation to the State Key Lab of Advanced Technology for Materials Synthesis and Processing (Wuhan University of Technology) for financial support. F.V. acknowledges the Chinese Central Government for an "Expert of the State" position in the program of "Thousand talents". H.D.V. appreciates the financial support from the Mexican Council of Science and Technology (CONACYT, Ph. D. grant). M.V. is a post-doctoral fellow with the FWO. Y.R. is indebted to the Marie Curie ITN with grant agreement number 238679. Computational resources and services were provided by Ghent University (Stevin Super-computer Infrastructure). We would also like to thank Prof. Dr ir. Veronique Van Speybroeck for reading the manuscript and scientific discussions.

Notes and references

1 M. Meldal and C. W. Tornøe, *Chem. Rev.*, 1980, **108**, 2952–3015; G.-C. Kuang, *et al.*, *J. Am. Chem. Soc.*, 2011, **133**, 13984–14001.
2 B. J. Adzima, C. J. Kloxin, C. A. DeForest, K. S. Anseth and C. N. Bowman, *Nat. Chem.*, 2011, **3**, 256–259.
3 S. Diez-Gonzalez, A. Correa, L. Cavallo and S. P. Nolan, *Chem. – Eur. J.*, 2006, **12**, 7558–7564.
4 S. Diez-Gonzalez and S. P. Nolan, *Angew. Chem., Int. Ed.*, 2008, **47**, 8881–8884.
5 C. Nolte, P. Mayer and B. F. Straub, *Angew. Chem., Int. Ed.*, 2007, **46**, 2101–2103.
6 Q. Wang, T. R. Chan, R. Hilgraf, V. V. Flokin, K. B. Sharpless and M. G. Finn, *J. Am. Chem. Soc.*, 2003, **125**, 3192–3193.
7 T. R. Chan and V. V. Fokin, *QSAR Comb. Sci.*, 2007, **26**, 1274–1279.
8 C. Gaulier, A. Hospital, B. Legeret, A. F. Delmas, V. Aucagne, F. Cisnetti and A. Gautier, *Chem. Commun.*, 2012, **48**, 4005–4007.
9 E. R. Stadtman, *Free Radical Res.*, 2006, **40**, 1250–1258.
10 E. R. Stadtman and B. S. Berlett, *Drug Metab. Rev.*, 1998, **30**, 225–243.
11 A. Kumar, K. Li and C. Cai, *Chem. Commun.*, 2011, **47**, 3186–3188.
12 R. V. Talanian, C. J. McKnight and P. S. Kim, *Science*, 1990, **249**, 769–771.
13 T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853–2855.
14 T. E. Ellenberger, C. J. Brandl, K. Struhl and S. C. Harrison, *Cell*, 1992, **71**, 1223–1237.
15 R. V. Talanian, C. J. McKnight, R. Rutkowski and P. S. Kim, *Biochemistry*, 1993, **31**, 6871–6875.
16 M. S. Sanford, J. A. Love and R. H. Grubbs, *J. Am. Chem. Soc.*, 2001, **123**, 6543–6554.
17 S. Monsaert, A. M. Lozano-Vila, R. Drozdzak, P. Van Der Voort and F. Verpoort, *Chem. Soc. Rev.*, 2009, **38**, 3360–3372.
18 A. M. Lozano-Vila, S. Monsaert, A. Bajek and F. Verpoort, *Chem. Rev.*, 2010, **110**, 4865–4909.
19 H. Díaz-Velazquez and F. Verpoort, *Chem. Soc. Rev.*, 2012, **41**, 7042–7060.
20 F. B. Hamad, T.-L. Sun, S.-Q. Xiao and F. Verpoort, *Coord. Chem. Rev.*, 2013, **257**, 2274–2292.
21 A. E. Cohrt, J. F. Jensen and T. E. Nielsen, *Org. Lett.*, 2010, **12**(23), 5414–5417.
22 T. Jin, S. Kamijo and Y. Yamamoto, *Eur. J. Org. Chem.*, 2004, 3789–3791.
23 J. C. Loren, A. Krasiński, V. V. Fokin and K. B. Sharpless, *Synlett*, 2005, 2847–2850.
24 L.-H. Lu, J.-H. Wu and C.-H. Yang, *J. Chin. Chem. Soc.*, 2008, **55**, 414–417.
25 W. Zhang, C. Kuang and Q. Yang, *Synthesis*, 2010, 283–287.
26 J. Barluenga, C. Valdés, G. Beltrán, M. Escribano and F. Aznar, *Angew. Chem., Int. Ed.*, 2006, **45**, 6893–6896.
27 L.-Y. Wu, Y.-X. Xie, Z.-S. Chen, Y.-N. Niu and Y.-M. liang, *Synlett*, 2009, **9**, 1453–1456.
28 Y. Jiang, C. Kuang and Q. Yang, *Synlett*, 2009, **19**, 3163–3166.
29 S. H. Cho and S. Chang, *Angew. Chem., Int. Ed.*, 2007, **46**, 1897–1900.
30 F. Zsila, Z. Bikadi, D. Malik, P. Hari, I. Pechand, A. Berces and E. Hazari, *Bioinformatics*, 2011, **27**(13), 1806–1813.



Food Additives & Contaminants: Part A
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tfac20

An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies

Melanie Sanders[a], Yirong Guo[ab], Abhishek Iyer[c], Yara Ruiz García[c], Anastasia Galvita[a], Arne Heyerick[d], Dieter Deforce[d], Martijn D.P. Risseeuw[e], Serge Van Calenbergh[e], Marc Bracke[f], Sergei Eremin[g], Annemieke Madder[c] & Sarah De Saeger[a]

[a] Laboratory of Food Analysis, Ghent University, Ghent, Belgium

[b] Institute of Pesticide and Environmental Toxicology, Zhejiang University, Hangzhou, China

[c] Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Belgium

[d] Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

[e] Laboratory for Medicinal Chemistry, Ghent University, Ghent, Belgium

[f] Laboratory of Experimental Cancer Research, Ghent University, Ghent, Belgium

[g] Department of Chemical Enzymology, Moscow State University, Moscow, Russia

Published online: 18 Sep 2014.

Food Additives & Contaminants: Part A, 2014
http://dx.doi.org/10.1080/19440049.2014.955887

An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies

Melanie Sanders[a]*, Yirong Guo[a,b], Abhishek Iyer[c], Yara Ruiz García[c], Anastasia Galvita[a], Arne Heyerick[d], Dieter Deforce[d], Martijn D.P. Risseeuw[e], Serge Van Calenbergh[e], Marc Bracke[f], Sergei Eremin[g], Annemieke Madder[c] and Sarah De Saeger[a]

[a]*Laboratory of Food Analysis, Ghent University, Ghent, Belgium;* [b]*Institute of Pesticide and Environmental Toxicology, Zhejiang University, Hangzhou, China;* [c]*Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Belgium;* [d]*Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium;* [e]*Laboratory for Medicinal Chemistry, Ghent University, Ghent, Belgium;* [f]*Laboratory of Experimental Cancer Research, Ghent University, Ghent, Belgium;* [g]*Department of Chemical Enzymology, Moscow State University, Moscow, Russia*

(*Received 5 June 2014; accepted 13 August 2014*)

An immunogen synthesis strategy was designed to develop anti-deoxynivalenol (DON) monoclonal antibodies with low cross-reactivity against structurally similar trichothecenes. A total of eight different DON immunogens were synthesised, differing in the type and position of the linker on the DON molecule. After immunisation, antisera from mice immunised with different DON immunogens were checked for the presence of relevant antibodies. Then, both homologous and heterologous enzyme-linked immunosorbent assays (ELISAs) were performed for hybridoma screening. Finally, three monoclonal antibodies against DON and its analogues were generated. In addition, monoclonal antibody 13H1 could recognise DON and its analogues in the order of HT-2 toxin > 15-acetyldeoxynivalenol (15-ADON) > DON, with IC_{50} ranging from 1.14 to 2.13 μg ml^{-1}. Another monoclonal antibody 10H10 manifested relatively close sensitivities to DON, 3-acetyldeoxynivalenol (3-ADON) and 15-ADON, with IC_{50} values of 22, 15 and 34 ng ml^{-1}, respectively. Using an indirect ELISA format decreases the 10H10 sensitivity to 15-ADON with 92%. A third monoclonal antibody 2A9 showed to be very specific and sensitive to 3-ADON, with IC_{50} of 0.38 ng ml^{-1}. Using both 2A9 and 10H10 monoclonal antibodies allows determining sole DON contamination.

Keywords: deoxynivalenol; monoclonal antibodies; immunogens; ELISA; cross-reactivity

Introduction

Trichothecene mycotoxins are a group of naturally occurring secondary metabolites produced by *Fusarium* species, in particular *F. graminearum* and *F. culmorum*. Among the 150 related trichothecenes, deoxynivalenol (DON) (Figure 1) is of special importance as it is formed in the field prior to harvest and because its occurrence cannot be completely avoided by plant production-minimising strategies due to the high impact of weather conditions. Especially wheat, triticale and maize are vulnerable for *Fusarium* infection and subsequent DON production (Maragos et al. 2006; Döll & Dänicke 2011).

DON can cause disease in several animal species, especially in swine, and causes symptoms including reduced feed consumption, abdominal distress, increased salivation, malaise, diarrhoea, anorexia, leucopenia, haemorrhage, shock and death in extremely high DON doses (Maragos & McCormick 2000; Pestka 2007). Besides these effects, DON is also generally considered to be a potent inhibitor of protein and DNA synthesis and is known to be immunosuppressive (Krska et al. 2001; Danicke et al. 2006). Highly dividing cells, such as intestinal epithelial cells known to ensure a proper barrier function, are especially sensitive to trichothecenes. Exposure to DON can lead to a decrease in absorption of nutrients and a decrease in cell proliferation and consequently cell differentiation (Pinton et al. 2012).

Next to DON two acetylated derivatives are known to be coproduced, namely 3-acetyldeoxynivalenol (3-ADON) and 15-acetyldeoxynivalenol (15-ADON). JECFA stated that the acetylated DON levels were less than 10% of the total DON found in cereal grains, whereas De Boevre et al. (2012) found higher 3-ADON and 15-ADON co-contamination for several food and feed samples. In 2010 JECFA considered the toxicity of the acetylated derivatives equal to DON, but a recent study suggested that higher toxicity of 15-ADON should be taken into account. Another DON masked form detected in cereals and beers, namely DON-3-glucoside (DON-3-G), also shows a non-negligible contribution to the overall DON contamination (Berthiller et al. 2009; JECFA 2011; De Boevre et al. 2012; Pinton et al. 2012).

*Corresponding author. Email: melanie.sanders@ugent.be

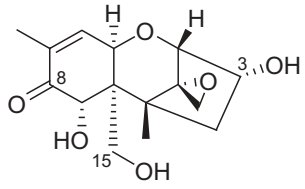


Figure 1. Chemical structure of deoxynivalenol.

Because of the economic importance of this toxin and from a food safety perspective, a variety of analytical techniques have been developed for the detection of DON and related trichothecene mycotoxins in food and feed. Most commonly used techniques include TLC, GC, HPLC and immunological approaches such as radioimmunoassay and ELISA. Recently, ELISA and ELISA-based procedures have gained acceptance as they offer the advantage of specificity, sensitivity, simplicity and rapidity, which are of importance for routine testing of mycotoxins (Maragos & McCormick 2000; Maragos et al. 2006, 2012). The specificity and sensitivity are dependent on the antibody used in the ELISA assay. Several antibodies for DON have been reported and all show high cross-reactivity against the acetylated derivatives 3-ADON and 15-ADON, which reduces their specific character. Sensitivity values of antibodies, measured by the IC_{50} value, range from approximately 20 to 3 ng ml^{-1} (Casale et al. 1988; Mills et al. 1990; Usleber et al. 1991; Sinha et al. 1995; Maragos & McCormick 2000).

The objective of this study is to improve the specificity of the ELISA assay by developing highly specific monoclonal anti-DON antibodies. The high affinity of the monoclonal antibody to 3-ADON or 15-ADON likely derives from the chemistry used to prepare the immunogen. The ester linkage of 3-ADON or 15-ADON may resemble the linkage of the DON-carrier protein immunogen. Therefore, it was decided to use different procedures to make linkers with varying length and chemical structure between DON and the carrier protein (Figure 2).

Materials and methods

Reagents

DON, 3-ADON and 15-ADON standards were obtained from Fermentek (Jerusalem, Israel). HT-2 toxin was purchased from Sigma-Aldrich (Bornem, Belgium).

Colorburst™ blue 3,3′,5,5′-tetramethylbenzidine (TMB) substrate solution containing hydrogen peroxide was supplied by Alerchek (Springvale, ME, USA). Rabbit anti-mouse immunoglobulins (anti-mouse IgG secondary antibody; protein concentration of 2.5 g l^{-1}) were purchased from DakoCytomation (Glostrup, Denmark). N, N′-carbonyldiimidazole, cyanuric chloride (CC), N,N′-diisopropylethylamine, glutaric anhydride, 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC), 1-butane boronic acid, sodium tetraborate, carboxymethoxylamine hemihydrochloride (CMO), γ-aminobutyric acid, sulfo-N-hydroxysuccinimide (NHS), tetrahydrofuran (THF), acetic acid, potassium carbonate, copper (II) sulfate, sodium ascorbate, 4,7-diphenyl-1,10-bathophenanthroline disulfonic acid disodium salt, arginine, glycine, aspartic acid, D-phenylalanine, lysine, trifluoroacetic acid (TFA), triisopropylsilane (TIS), methyl tertiary butyl ether (MTBE), polyethylene glycol (bovine serum albumin (BSA), ovalbumin (OVA), horseradish peroxidase (HRP), rabbit anti-mouse IgG secondary antibody labelled with horseradish peroxidase (Sec Ab-HRP), PBS (0.01 M, pH 7.4) tablet, carbonate-buffered saline (CBS, 0.05 M, pH 9.6) capsule, complete and incomplete Freund's adjuvants, Tween 20 and skim milk powder were obtained from Sigma-Aldrich (Bornem, Belgium). O-propargyl-hydroxylamine hydrochloride was purchased from Focus Synthesis (San Diego, CA, USA). Imidazole-1-sulfonylazide hydrochloride was kindly provided by the Laboratory of Medicinal Chemistry (Ghent, Belgium). The Oasis HLB cartridges were obtained from Waters (Zellik, Belgium). Chlorotrityl chloride resin was purchased from Merck (Darmstadt, Germany). Deionised water was purified by a Millipore Milli-Q system (Brussels, Belgium). Other chemicals and solvents were of analytical grade.

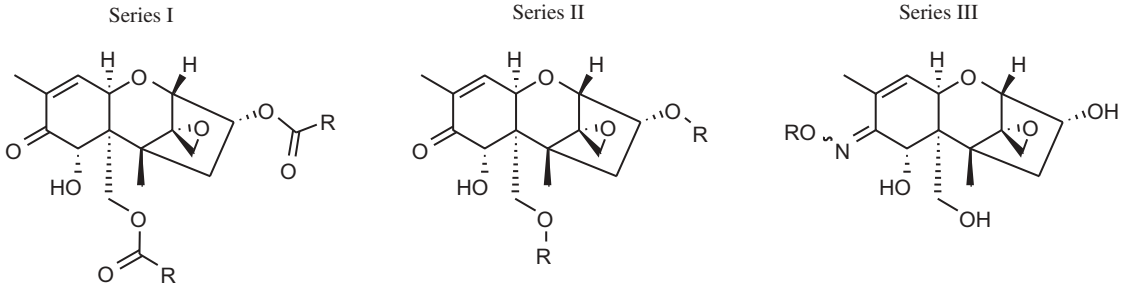


Figure 2. Three series of synthesised DON immunogens.

Nunc-Immuno™ F96-well microplates and Nunclon™ cell culture plates were from Nalge Nunc International (Roskilde, Denmark). Protein concentrators (9K MWCO, 20 ml) were purchased from Thermo Scientific (Rockford, USA).

Preparation of DON immunogens

DON is a hapten due to its small size and therefore cannot elicit an immune response. Consequently, for the synthesis of immunogens, DON was coupled to a protein via a linker to increase the molecular weight. In search for specific anti-DON antibodies, different synthetic strategies were followed. Firstly, DON was coupled to a carrier protein via the C3 and/or C15 position using a linker with a carboxyl function (Series I, Figure 2). To minimise the chance of cross-reacting monoclonal antibodies, a second DON series of immunogens was generated by coupling the carrier protein via a linker on C3 and/or C15 without a carboxyl function (Series II, Figure 2). To reduce even further the possibility of cross-reacting monoclonal antibodies the linkage to the carrier protein was introduced through reaction with the C8 carbonyl function of DON, delivering a last series of immunogens (Series III, Figure 2). The used carrier proteins were bovine serum albumin (BSA) for the immunogen and ovalbumin (OVA) or horseradish peroxidase (HRP) for the coating/competitive antigen in indirect or direct enzyme-linked immunosorbent assay (ELISA), respectively. The synthetic pathways underneath are described for BSA. The basic chemical structure of the three different series of immunogens is illustrated in Figure 2.

The first immunogen of the series I is DON-3,15-hemiglutaryl (HG)-BSA (Figure 3A) and was formed by the reaction of DON with glutaric anhydride in dry pyridine. Concisely, to a solution of 5 mg of DON in 500 µl of dry pyridine, 100 mg of glutaric anhydride was added and the solution was heated at 100°C for 8 h. After evaporation, the residue was dissolved in 7.5 ml of chloroform and washed three times with 5 ml of 0.1 M HCl. A purification step was performed using Oasis HLB® cartridges by a method derived from De Smet et al. (2010). Briefly, 5 µg of the crude DON-HG conjugate was dissolved in 1 ml of methanol–water (10:90, v/v) pH 2.3 and loaded onto a preconditioned column. After loading, the sorbent was dried for 15 s by applying vacuum. DON-HG was eluted by washing the column with 1 ml of methanol–water (55:45, v/v) followed by 1 ml of methanol. The sorbent was dried for 15 s by applying a vacuum. After purity determination by time-of-flight (TOF) MS, an aliquot of 1 mg of the synthesised DON-HG (1.92 µmol) was dissolved in 0.5 ml PBS using ultrasonication and transferred to 30 mg BSA (0.45 µmol) in 1 ml of PBS. Then 20 mg of EDC (0.10 µmol) was added to the above mixture and incubated for 5 h at RT with stirring (Mills et al. 1990; Usleber et al. 1991; Kohno et al. 2003; De Smet et al. 2010).

To obtain DON with one HG linker, namely DON-3-HG-BSA (Figure 3B), prior protection of the 7-OH and 15-OH groups as a cyclic boronate ester allowed one to use similar reaction conditions as for DON-3,15-HG-BSA (Casale et al. 1988). Additionally in the first series of immunogens, a common DON-BSA carbamate conjugate (Figure 3C and 3D) was synthesised by the N,N′-carbonyldiimidazole (CDI) coupling reaction, adopted from the published literature (Xiao et al. 1995; Maragos & McCormick 2000).

Within-series II, DON-CC-BSA (Figure 3E and 3F) was synthesised by first dissolving 1 mg of DON (3.374×10^{-3} mmol) in 480 µl of cold acetonitrile. A solution of 620 µg of cyanuric chloride (3.362×10^{-3} mmol) in 1.24 ml of acetonitrile was prepared and cooled until –20°C. The DON solution was added over 1 h to the vigorously stirred solution of cyanuric chloride. A solution of 870 µg N,N′-diisopropylethylamine in 260 µl of cold acetone was added to the solution of DON and cyanuric chloride and mixed for 5 h at 55–60°C followed by mixing for 16 h at RT. For coupling to BSA, a solution of 1830 µg of BSA (0.0270 µmol) in 1090 µl of 0.1 M sodium tetraborate (pH 9.2) was prepared. The solution was cooled to 4°C and 3.5 mg of DON, a 10-fold molar excess with respect to the available amino groups was added and the resulting mixture was stirred for 1 h at 4°C (Abuchowski et al. 1977; Abuknesha & Griffith 2005).

In an effort to reduce the cross-reactivity against 3-ADON and 15-ADON, we decided to use the C8 carbonyl function of DON to attach an appropriate linker via an oxime moiety (Series III). Hereto a carboxymethyloxime (CMO) strategy was applied (Figure 3G) (Burkin et al. 2000). To investigate the effect of the length of the linker between DON and the protein and with the aim of lowering the immune response against the linker itself, we also introduced a carboxypropyl imine (CPI) linker at the C8 position of DON (Figure 3H). For this purpose, 0.5 mg of DON (1.7 µmol) was reacted with 750 µg of CMO (7 µmol) in 500 µl of dry pyridine for the DON-CMO synthesis and with 1.6 mg of γ-aminobutyric acid (15.5 µmol) in 1 ml dimethylformamide (DMF) for the DON-CPI synthesis. After overnight reaction at RT, the reaction mixture was concentrated and the residue redissolved in 200 µl of water/DMSO (1:1.5, v/v). Next, 2 mg of sulfo-NHS and 4 mg of EDC were added together with 1.775 mg of BSA and the volume was adjusted to 1 ml by further dilution with water. The reactions were performed for 2 h at RT followed by washing and purification as described for previous immunogen synthesis.

Finally, click chemistry was adopted for linking C8 of DON without a carboxyl function. Following a modified procedure of Ikuina et al. (2003), DON was condensed to

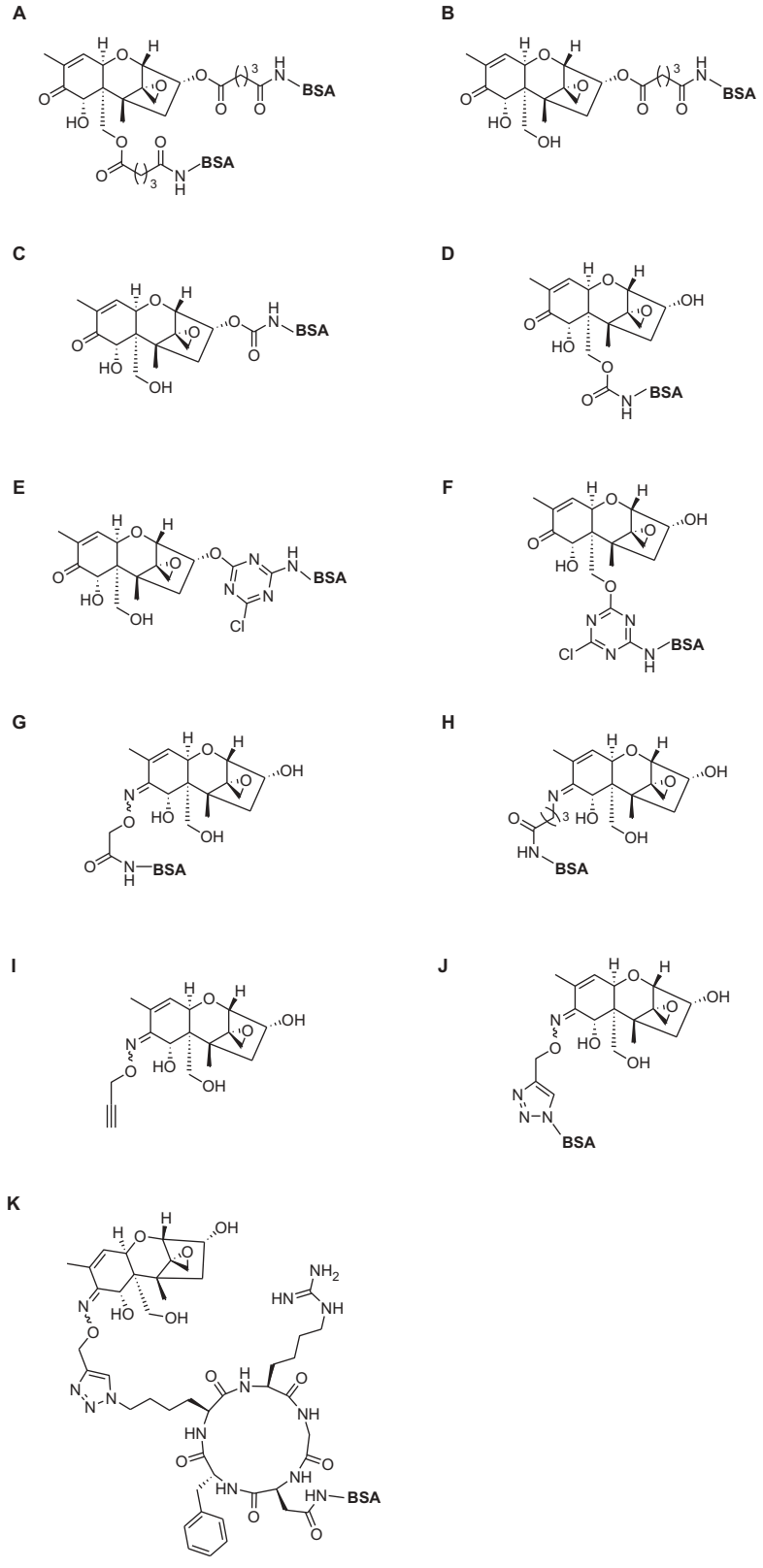


Figure 3. Overview of the deoxynivalenol immunogens (only the amino group of the protein is presented): A, DON-3,15-HG protein; B, DON-3-HG protein; C, DON-3 protein; D, DON-15 protein; E, DON-3-CC protein; F, DON-15-CC protein; G, DON-CMO protein; H, DON-CPI protein; I, DON-oxime; J, DON-azido protein; K, DON-cyclic peptide protein.

O-propargy-hydroxylamine hydrochloride (Figure 3I). Briefly, 2 mg of DON (6.76 µmol) was reacted with 1.81 mg of O-propargyl-hydroxylamine hydrochloride (25.52 µmol) in the presence of THF/acetic acid (1:1, v/v) at 40°C. After 2 h, an additional 170 µg of O-propargyl-hydroxylamine hydrochloride was added and the mixture was stirred for another 4.5 h to obtain DON-oxime. For the synthesis of azido-BSA, 10.686 mg of BSA (0.16 µmol) was dissolved in water together with 4 mg of potassium carbonate and 0.5 mg of copper (II) sulfate. Then, 1 mg of imidazole-1-sulfonylazide hydrochloride (4.79 µmol) was added and the reaction was agitated overnight. For the click reaction, 0.5 mg of DON-oxime (1.43 µmol) was transferred to an aqueous solution of azido-BSA (1.42 mg) containing 25 µg copper (II) sulfate (10 mM), 20 µg sodium ascorbate (10 mM) and 54 µg 4,7-diphenyl-1,10-bathophenanthroline disulfonic acid disodium salt (10 mM). The reaction went on for 14 h at RT (Figure 3J) (Ikuina et al. 2003; Van Dongen et al. 2009; Horak et al. 2010).

The click chemistry immunogen product was further expanded by the introduction of a N-azido cyclic peptide (630.3 g mol^{-1}) (Figure 3K). The cyclic peptide increases the distance between DON and BSA and renders the DON molecule more available for the immune system. Through solid-phase peptide synthesis using 2-chlorotrityl chloride resin and a Fmoc/*t*Bu protection scheme, the amino acids glycine, aspartic acid, D-phenylalanine, lysine and arginine were successively coupled to each other. After mild acid cleavage of the peptide from the resin, cyclisation of the peptide was performed followed by treatment with TFA–TIS–H_2O (95:2.5:2.5, v/v/v) to remove the protecting groups in solution. The peptide was then precipitated in MTBE–hexane (1:1, v/v) and redissolved in methanol for a diazo transfer. For the click chemistry, 1 mg of DON-oxime (2.84 µmol) was reacted with 1.78 mg of cyclic peptide (2.84 µmol) under the previously mentioned conditions. Under CDI reaction conditions, the obtained DON-cyclic peptide (2.84 µmol) was coupled to 3.9 mg of BSA in water–DMSO (1/5, v/v) for 3 h at RT (Dai et al. 2000; Dijkgraaf et al. 2007).

All DON immunogens were dialysed against 4 L PBS to remove low molecular weight substances and concentrated using Pierce Concentrator columns 20 ml/9K MWCO.

Characterisation of immunogens

The successful synthesis of the DON-3,15-HG-linker was confirmed by mass spectrometry LCT Premier XETM TOF (Waters, Milford, MA, USA) equipped with a standard electrospray ionisation and modular LockSpray TM interface in the positive and negative electrospray ionisation ($ESI^{+/-}$) mode. The purified DON-3,15-HG mixture was infused in acetonitrile–water (1:1, v/v) at 10 µl min^{-1}. The purity of the final product was assessed by HPLC and photo diode array (PDA) detection (190–400 nm) using a Phenomenex Luna 2.5 mm C_{18} (2)-HST column. A mobile phase consisting of eluents A (water, 0.1% formic acid) and B (acetonitrile, 0.1% formic acid) was used at a flow rate of 0.4 ml min^{-1}. A linear gradient of 10–100% solvent B was applied over 9 min. Other mass measurements of DON-linker syntheses were performed by direct injection of the reaction product into the Quattro Premier XE mass spectrometer (Waters, Milford, MA, USA) using the $ESI^{+/-}$ mode. Masslynx version 4.1 was used for data acquisition.

After coupling to the protein, the immunogen concentration was determined based on the amount of protein coupled by the use of the NanoDrop 2000c (Thermo Scientific, Rockford, USA) and the immunogen was characterised by the use of indirect competitive ELISA. All incubations except for the first coating step were carried out at 37°C. After each incubation, the plates were washed three times (300 µl/well) with PBST (PBS containing 0.05% Tween 20). High-binding Nunc-Immuno^{TM} F96-well microplates were coated with 10 µg ml^{-1} of the synthesised conjugate in CBS (100 µl/well). After incubation at 4°C overnight, the plates were blocked with 2% skimmed milk in PBS (300 µl/well) for 30 min. Standard solutions of DON and PBS control were added (50 µl/well), followed by adding 50 µl/well of diluted DON monoclonal antibodies (clone 4 or 22) in PBS kindly provided by C. Maragos (USDA). After shaking and incubation for 1 h, 100 µl/well of Sec Ab-HRP was added and incubated for another 1 h. Then, 100 µl/well of TMB substrate solution was added. The reaction was stopped after 15 min with 2 M sulphuric acid (50 µl/well), and the absorbance at 450 nm was measured by a Bio-Rad 550 microplate reader (Richmond, CA, USA).

Immunisations

To obtain antibody-producing B-lymphocytes against DON, 6–8-week-old Balb/C female mice (ethical approval according to ethical commission for animals (ECD) 10/08) were subjected to an injection with the different DON immunogens emulsified with complete or incomplete Freund's adjuvant. For each group of mice, 100 µg of the DON immunogen was administered. Once the mice had reached a sufficient titre, cell fusion of mouse spleen cells and myeloma cells (NSO cells) was performed. Polyethylene glycol (PEG) 1500 was added as fusing reagent and hypoxanthine, aminopterine and thymidine (HAT) for the selection of the fused cells. The cells were distributed into 96-well culture plates and cultured in a humidified, 37°C, 5% CO_2 incubator. Culture supernatant was screened by indirect ELISA to determine the positive hybridomas producing antibody against DON. These hybridomas were further screened for the production of the target antibody and subcloned by limiting dilution.

Characterisation of monoclonal antibodies

Checkerboard assays, in which antibodies were titrated against various amounts of coating antigens, were conducted to select appropriate working concentrations for evaluation of assay sensitivities to DON. Standard competitive curves were obtained by plotting relative absorbance (ratio of absorbance measured at the standard concentration and zero concentration: $B/B_0 \times 100\%$) against the logarithm of analyte concentration. IC_{50} values (i.e., analyte concentrations at which the maximum absorbances were inhibited by 50%) were determined to assess the assay sensitivity. To evaluate the specificity or assay selectivity of the antibody, a set of DON analogues were utilised to perform cross-reactivity studies. The IC_{50} of each tested compound was based on its corresponding competitive curve. Cross-reactivity (CR) values were calculated according to the following equation: CR (%) = [IC_{50} (DON)/IC_{50} (analogue)] × 100%, where CR values were calculated using IC_{50} values with units of μg ml^{-1}.

Results and discussion

Characterisation of immunogens

When the desired synthesis products were obtained based on the exact mass measurements performed by direct injection into the Quattro Premier XE mass spectrometer or the LCT Premier XETM TOF mass spectrometer, further coupling to BSA and OVA or HRP was performed. The synthesised immunogens were characterised by competitive ELISA using reference DON monoclonal antibodies (clone 4 or 22, USDA). If the B_0 value was equal to or higher than 1 and decreasing B values were obtained when using increasing standard DON concentrations, the testing conjugates were confirmed and used for immunisation or further ELISA experiments.

Determination of antisera titres by using indirect ELISA

Antisera titres were determined by an indirect homologous ELISA using the immunising haptens coupled to OVA instead of BSA as coating antigen (Guo et al. 2014). By using the same format, the cross-reactivity against OVA was determined as well. The best antisera titres were obtained for the mice injected with DON-BSA and DON-CC-BSA. For DON-3,15-HG-BSA, high antisera titres were found, but the cross-reactivity against OVA was high. It was concluded that antibodies were probably formed against the protein instead of the target DON. This could be explained by the presence of two linkers on the DON molecule, which makes it less free for activation of the immune system of the mouse. For DON-CMO-BSA, DON-CPI-BSA, DON-azido-BSA and DON-cyclic peptide-BSA, a relatively lower titre, but no cross-reactivity against OVA was determined. When the ELISA response reached a plateau phase using a high serum dilution, the mice were sacrificed by cervical dislocation.

Hybridoma selection and subcloning

By screening and subcloning of the hybridomas, finally, a total of three different anti-DON monoclonal antibodies were obtained. Corresponding to the derived antisera titres, one DON-CC-BSA mouse produced the 13H1 monoclonal antibody, while the 10H10 and 2A9 monoclonal antibodies were produced by two different DON-BSA mice. Previous statement about the immunogenic importance of the C8 carbonyl function in the DON molecule was confirmed as no anti-DON monoclonal antibodies were derived using immunogens with a linker coupled to the C8 carbonyl function (Usleber et al. 1991).

Characterisation of monoclonal antibodies

The characterisation of monoclonal antibodies 13H1 and 10H10 by direct ELISA is described in detail by Guo et al. (2014). The 13H1 and 10H10 monoclonal antibody showed sensitivity values of 2.13 and 0.022 μg ml^{-1}, respectively. For the 13H1 monoclonal antibody the highest cross-reactivity was observed for 15-ADON (CR = 131%) and HT-2 toxin (CR = 187%). Monoclonal antibody 10H10 showed the highest cross-reactivity for 3-ADON (CR = 147%) and 15-ADON (CR = 65%). Characterisation of 10H10 was also repeated by indirect competitive ELISA using DON-OVA (4 μg ml^{-1}) and DON-CC-OVA (4 μg ml^{-1}) coating and DON, 3-ADON and 15-ADON for competition. This comparison between direct and indirect ELISA is illustrated in Table 1. The sensitivity of the antibody towards DON and 3-ADON measured by indirect ELISA is approximately half (DON-OVA coating) and 1/5 (DON-CC-OVA coating) of the value measured by direct ELISA. For 15-ADON, the sensitivity decreased at least ten times when using indirect ELISA. When looking to the CR_{molar} values for direct ELISA and indirect ELISA with DON-OVA coating, the same cross-reactivity is seen for 3-ADON, but the cross-reactivity towards 15-ADON lowered 6.7 times when using indirect ELISA. It can even be concluded that the 10H10 antibody shows only cross-reactivity to 3-ADON when using indirect ELISA with DON-OVA coating. When using an indirect ELISA format with DON-CC-OVA coating, the cross-reactivity against 3-ADON increased three times and the cross-reactivity against 15-ADON decreased four times compared to the direct ELISA format. So, the characterisation of the monoclonal antibody depends on the type of ELISA and the coating antigen used.

Characterisation of monoclonal antibody 2A9 was performed using direct ELISA and the standard curve using 3-ADON standard for competition is shown in Figure 4. The curve represents the relative absorbance

Table 1. Comparison between direct and indirect ELISA for 10H10 antibody characterisation.

			IC_{50} (μg ml^{-1})	CR (%)	IC_{50} molar (nmol ml^{-1})	CR molar (%)
Direct ELISA		DON	0.022	100	0.074	100
		3-ADON	0.015	147	0.044	167
		15-ADON	0.034	65	0.101	74
Indirect ELISA	DON-OVA	DON	0.040	100	0.135	100
		3-ADON	0.028	142	0.083	162
		15-ADON	0.421	9	1.250	11
	DON-CC-OVA	DON	0.113	100	0.383	100
		3-ADON	0.024	480	0.070	547
		15-ADON	0.637	18	1.885	20

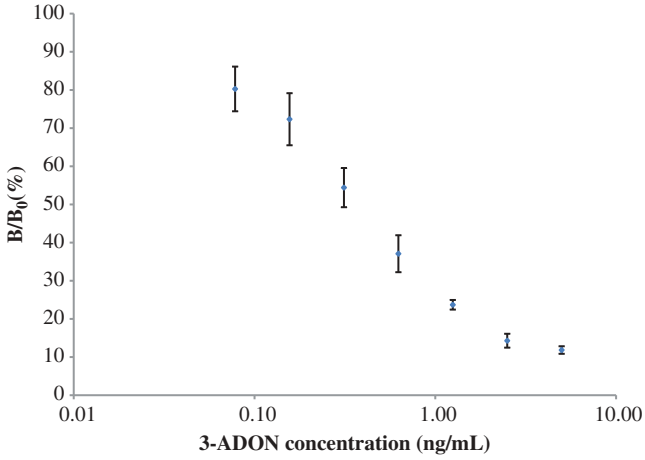


Figure 4. Standard curve of clone 2A9 for 3-ADON by direct competitive ELISA (n = 3).

values (B/B_0) of the ELISA experiment performed on 3 consecutive days. A very sensitive anti-3-ADON antibody was developed with an IC_{50} value of 0.38 ng ml^{-1}. Setting the monoclonal antibody activity for 3-ADON as 100%, the cross-reactivity values for DON, 15-ADON and DON-3-G (Table 2) were determined as 0.188%, 0.088% and 1.498% and can be considered as negligible. In comparison with other previously reported monoclonal antibodies for 3-ADON, the newly developed clone 2A9 is likely the most sensitive and specific antibody to 3-ADON (Casale et al. 1988; Maragos & McCormick 2000; Maragos et al. 2006; Baumgartner et al. 2010).

Conclusions

By using a three-series synthesis strategy for the development of specific monoclonal antibodies against DON, previous statements were confirmed and new conclusions could be made. The place of the linker on the DON molecule is of importance for the immunogenic response. When a linker is positioned on the C3 or C15 of DON for the synthesis of an immunogen, the produced antibody shows higher cross-reactivity against 3-ADON or 15-ADON, respectively. When coupling a linker to the carbonyl C8 of DON, no high anti-DON immune response was observed. This emphasises the immunogenic importance of the C8 carbonyl function. The size of the linker does not seem to influence the immunogenicity of the injected conjugate. The kind of linker has an influence on the electronic configuration of the DON molecule and therefore on the antigenic determinant of DON which is available for the immune system of the mouse.

Three different monoclonal antibodies were developed. One of them (2A9) can be called an anti-3-ADON monoclonal antibody, because of its very sensitive and specific characteristics towards 3-ADON. The second 10H10 monoclonal antibody is a broad specific antibody against DON and its acetylated derivatives 3-ADON and

Table 2. Comparison of cross-reactivity results between clone 2A9 and other DON-MAbs.[a]

	Clone 2A9 direct ELISA		Mab 1–6.2.6 (1) indirect ELISA		USDA Clone 22 (4) direct ELISA	
Compound	IC_{50} (ng ml^{-1})	CR (%)	IC_{50} (ng ml^{-1})	CR (%)	IC_{50} (ng ml^{-1})	CR (%)
3-ADON	0.38	100	1.70	100	2.88	100
DON	202.14	0.188	15.80	10.7	18.20	15.8
15-ADON	431.79	0.088	68.90	2.4	558.00	0.52
DON-3-G	25.36	1.498	n.a.	–	n.a.	–

Notes: [a]Herein, CR (%) = [IC_{50} (3-ADON)/IC_{50} (analogue)] × 100%.
For comparison, the results from reference (4) by Maragos & McCormick (2000) were competitive direct ELISA for clone 22; the results from reference (1) by Maragos et al. (2006) were competitive indirect ELISA for Mab 1–6.2.6.

15-ADON when using it in a direct ELISA format. By using an indirect ELISA format with DON-OVA or DON-CC-OVA coating, the selectivity of this antibody changes resulting in only cross-reactivity against 3-ADON. When combining the 2A9 and 10H10 antibodies in an indirect ELISA format, it is possible to measure the sole DON contamination.

Therefore, we conclude that it is possible to produce specific antibodies against trichothecene mycotoxins such as 3-ADON by synthesising the proper immunogen with the suitable linker and screening the hybridomas carefully. By changing the ELISA format and/or coating antigen, it is possible to influence the selectivity and cross-reactivity of the monoclonal antibody.

Acknowledgements

This work was executed thanks to the contribution of Ateknea Solutions (Hungary), ULUND (Sweden), HGFA (Hungary), SEEDYZ (Greece), CESFAC (Spain), ASEMAC (Spain), Synagra (Belgium), EASRET (Greece), Impuls Ltd (Poland), OSV Srl (Italy), Dunagabona Ltd (Hungary), Dimitriaki S.A. (Greece), ETIA (France), Bioforum S. A. (Greece), and EST Ltd (UK).

Funding

Research leading to the results received funding from the European Community's Seventh Framework Program (FP7/2007-2013) in the frame of the MYCOHUNT project [grant agreement number 243633].

Supplemental data

Supplemental data for this article can be accessed online: http://dx.doi.org/10.1080/19440049.2014.955887

References

Abuchowski A, Van Es T, Palczuk N, Davis F. 1977. Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. J Biol Chem. 252:3578–3581.

Abuknesha R, Griffith H. 2005. Generation of antiserum to Irgarol 1051 and development of a sensitive enzyme immunoassay using a new heterologous hapten derivative. Anal Bioanal Chem. 381:233–243.

Baumgartner S, Führer M, Krska R. 2010. Comparison of monoclonal antibody performance characteristics for the detection of two representatives of A- and B-trichothecenes: T-2 toxin and deoxynivalenol. World Mycotoxin J. 3:233–238.

Berthiller F, Dall'asta C, Corradini R, Marchelli R, Sulyok M, Krska R, Adam G, Schuhmacher R. 2009. Occurrence of deoxynivalenol and its 3-β-D-glucoside in wheat and maize. Food Addit Contam: Part A. 26:507–511.

Burkin A, Kononenko G, Soboleva N, Zotova E. 2000. Preparation of conjugated antigens based on zearalenone carboxymethyloxime and their use in enzyme immunoassay. Appl Biochem Microbiol. 36:282–288.

Casale WL, Pestka JJ, Hart LP. 1988. Enzyme-linked immunosorbent assay employing monoclonal antibody specific for deoxynivalenol (vomitoxin) and several analogs. J Agr Food Chem. 36:663–668.

Dai XD, Su Z, Liu JO. 2000. An improved synthesis of a selective $\alpha_v\beta_3$-integrin antagonist cyclo(-RGDfK-). Tetrahedron Lett. 41:6295–6298.

Danicke S, Goyarts T, Doll S, Grove N, Spolders M, Flachowsky G. 2006. Effects of the Fusarium toxin deoxynivalenol on tissue protein synthesis in pigs. Toxicol Lett. 165:297–311.

De Boevre M, Di Mavungu JD, Landschoot S, Audenaert K, Eeckhout M, Maene P, Haesaert G, De Saeger S. 2012. Natural occurrence of mycotoxins and their masked forms in food and feed products. World Mycotoxin J. 5:207–219.

De Smet D, Monbaliu S, Dubruel P, Van Peteghem C, Schacht E, De Saeger S. 2010. Synthesis and application of a T-2 toxin imprinted polymer. J Chromatogr A. 1217:2879–2886.

Dijkgraaf I, Rijnders AY, Soede A, Dechesne AC, van Esse GW, Brouwer AJ, Corstens FHM, Boerman OC, Rijkers DTS, Liskamp RMJ. 2007. Synthesis of DOTA-conjugated multivalent cyclic-RGD peptide dendrimers via 1,3-dipolar cycloaddition and their biological evaluation: implications for tumor targeting and tumor imaging purposes. Org Biomol Chem. 5:935–944.

Döll S, Dänicke S. 2011. The Fusarium toxins deoxynivalenol (DON) and zearalenone (ZON) in animal feeding. Prev Vet Med. 102:132–145.

Guo Y, Sanders M, Galvita A, Heyerick A, Deforce D, Bracke M, Eremin S, De Saeger S. 2014. Heterologous screening of hybridomas for the development of broad-specific monoclonal antibodies against deoxynivalenol and its analogues. World Mycotoxin J. 7:257–265.

Horak J, Hofer S, Lindner W. 2010. Optimization of a ligand immobilization and azide group endcapping concept via "Click-Chemistry" for the preparation of adsorbents for antibody purification. J Chromatogr B Anal Technologies Biomed Life Sci. 878:3382–3394.

Ikuina Y, Amishiro N, Miyata M, Narumi H, Ogawa H, Akiyama T, Shiotsu Y, Akinaga S, Murakata C. 2003. Synthesis and antitumor activity of novel O-carbamoylmethyloxime derivatives of radicicol. J Med Chem. 46:2534–2541.

[JECFA] Joint FAO/WHO Expert Committee on Food Additives. 2011. Safety evaluation of certain contaminants in food. Prepared by the seventy-second meeting of the Joint FAO/WHO Expert Committee on Food Additives. *WHO Food Additives Series 63*.

Kohno H, Yoshizawa T, Fukugi M, Miyoshi M, Sakamoto C, Hata N, Kawamura O. 2003. Production and characterization of monoclonal antibodies against 3,4,15-triacetylnivalenol and 3,15-diacetyldeoxynivalenol. Food Agric Immunol. 15:243–254.

Krska R, Baumgartner S, Josephs R. 2001. The state-of-the-art in the analysis of type-A and -B trichothecene mycotoxins in cereals. Fresenius J Anal Chem. 371:285–299.

Maragos C, Busman M, Sugita-Konishi Y. 2006. Production and characterization of a monoclonal antibody that cross-reacts with the mycotoxins nivalenol and 4-deoxynivalenol. J Food Addit Contam. 23:816–825.

Maragos CM, Li L, Chen D. 2012. Production and characterization of a single chain variable fragment (scFv) against the mycotoxin deoxynivalenol. Food Agric Immunol. 23:51–67.

Maragos CM, McCormick SP. 2000. Monoclonal antibodies for the mycotoxins deoxynivalenol and 3-acetyl-deoxynivalenol. Food Agric Immunol. 12:181–192.

Mills C, Alcock S, Lee H, Morgan M. 1990. An enzyme-linked immunosorbent assay for deoxynivalenol in wheat, utilizing novel hapten derivatization procedures. Food Agric Immunol. 2:109–118.

Pestka JJ. 2007. Deoxynivalenol: toxicity, mechanisms and animal health risks. Anim Feed Sci Technol. 137:283–298.

Pinton P, Tsybulskyy D, Lucioli J, Laffitte J, Callu P, Lyazhri F, Grosjean F, Bracarense AP, Kolf-Clauw M, Oswald IP. 2012. Toxicity of deoxynivalenol and its acetylated derivatives on the intestine: differential effects on morphology, barrier function, tight junction proteins, and mitogen-activated protein kinases. Toxicol Sci. 130:180–190.

Sinha R, Savard M, Laur R. 1995. Production of monoclonal antibodies for the specific detection of deoxynivalenol and 15-acetyldeoxynivalenol by ELISA. J Agr Food Chem. 43:1740–1744.

Usleber E, Maertlbauer E, Dietrich R, Terplan G. 1991. Direct enzyme-linked immunosorbent assays for the detection of the 8-ketotrichothecene mycotoxins deoxynivalenol, 3-acetyldeoxynivalenol, and 15-acetyldeoxynivalenol in buffer solutions. J Agr Food Chem. 39:2091–2095.

van Dongen SFM, Teeuwen RLM, Nallani M, van Berkel SS, Cornelissen JJLM, Nolte RJM, van Hest JCM. 2009. Single-step azide introduction in proteins via an aqueous diazo transfer. Bioconjug Chem. 20:20–23.

Xiao H, Clarke JR, Marquardt RR, Frohlich AA. 1995. Improved methods for conjugating selected mycotoxins to carrier proteins and dextran for immunoassays. J Agr Food Chem. 43:2092–2097.

Annex II:

Articles to be submitted

DOI: 10.1002/anie.200((will be filled in by the editorial staff))

Sequence-Selective DNA Recognition and Enhanced Cellular Uptake by Peptide-Steroid Conjugates prepared through CuAAC

Yara Ruiz García[a], Abhishek Iyer[a], Dorien Van Lysebetten[a], Vladimir Pabón Martínez[c], Benoit Louage[b], Bruno De Geest[b], Edvard Smith[c] and Annemieke Madder[a]*

Abstract: *Several GCN4 Leucine Zipper Transcription Factor (TF) models have already been designed and synthesized. However, the synthetic routes for these constructs are tedious and difficult. Therefore there was need to find a convergent strategy to synthesize these constructs which could allow versatility without adding to the complexity of the synthesis. Through this work we have performed a substitution of the Leucine zipper domain of the protein by different derivatives of deoxycholic acid which have enabled us to synthesize mimics of the GCN4 TF using a one-step "click" reaction. In addition to achieving sequence specific dsDNA binding, we have investigated the potential of these compounds to enter cells through confocal microscopy and flow cytometry which have not yet been performed before for steroid-long peptide conjugated constructs nor for artificial DNA binders. We thus present a unique synthetic model of the bZIP TF which features sequence specific dsDNA binding properties as well as enhanced cell-uptake through active diffusion. Given the unique properties of deoxycholic acid and convergent nature of the synthesis, we believe this work is a keystone in the discovery of an ideal mimic for this* TF.

Gene expression at the transcriptional level is mainly regulated by proteins that bind DNA in a sequence-specific manner. These proteins, known as transcription factors, are responsible for controlling the transfer of genetic information from DNA to mRNA. Numerous genetic diseases find their origin in the anomalous regulation of gene expression during transcription. TFs are involved in the development of biological functions in the cell. More specifically, oncoproteins are transcription factors responsible for cell-growth proliferation and tumor formation. As a consequence of the specificity of these oncoproteins in the DNA sequence recognition during transcription, several approaches have been explored to develop inhibitors or modifiers of gene expression that

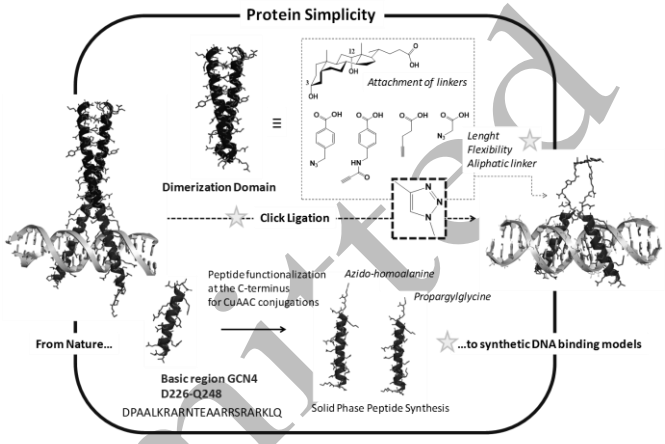


Figure 1. Artificial DNA binder design.

can prevent specific genes from being transcribed. This could ultimately result in interesting biomedical applications such as design of genome interfering agents or diagnostics. In addition, the lack of a general recognition code for the interaction between amino acid sequences within a protein and its specific DNA-binding site has promoted the study of the structure of TFs and its interaction with the DNA. We here present our efforts on mimicking the GCN4 bZIP TF by a synthetic approach. The broad range of chemical tools allowed us to simplify the structure of the protein and to explore factors that influence DNA recognition in DNA binder models. This coupled with cell uptake studies is a first attempt to understand how mimics of a protein behave in cellulo. Our target has been to develop new types of site-specific DNA binders which can recognize dsDNA by specific binding in the major groove and show enhanced cellular uptake by exploiting the unique properties of the steroid moiety. We have been inspired by the leucine zipper motifs (bZIP and b-HLH-ZIP), in which the basic DNA recognition region binds to the major groove as a dimer, inserting two α-helices held in the correct position by a dimerization domain[1]. The main residues of the protein involved in the DNA recognition comprise of the amino acids 226-248 of the basic region of the GCN4 protein. Previous models of such transcription factors have been synthesized by different research groups proposing different dimerization domains. The pioneer mimic was developed by the group of Kim[2], in which the dimerization domain was substituted by a simple disulfide bond. Following this successful idea, Morii, Schepartz, and Mascareñas have enforced the proof-of-concept by using other dimerizing moieties[3–7]. In order to mimic the DNA binding potential of this natural construct, we envisaged that the attachment of the basic region peptides to a rigid scaffold, a derivative of deoxycholic acid in this case, can allow selective recognition of DNA. Our previous work on Myc-Max b-HLH-ZIP and GCN4-bZIP proteins showed that these constructs have the potential to bind DNA. We illustrate the importance of the dimerizing unit for efficient recognition[8]. The specific choice of the steroid scaffold was inspired by the fact that it is cheap,

[a] Yara Ruiz García, Abhishek Iyer, Dorien Van Lysebetten, Prof. Dr. A. Madder. Organic and Biomimetic Chemistry Research Group. Department of Organic Chemistry, Ghent University. Krijgslaan 281 (S4), B-9000 Ghent (Belgium)
Fax: (+32) 9 264 4998. E-mail: annemieke.madder@ugent.be

[b] Benoit Louage, Prof. Dr. Ir. Bruno de Geest. Faculty of Pharmaceutical Sciences. Department of Pharmaceutics. Ghent University. Harelbekestraat 72, 9000 Ghent. Belgium. Tel.: +32 9 264 80 55. E-mail: br.degeest@ugent.be.

[c] Vladimir Pabón Martínez and Edvard Smith. Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. Tel.: 46-585-83651. Fax: 46-585-83650. E-mail: edvard.smith@ki.se

This work was supported by Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number (PITN-GA-2010; 238679).
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.((Please delete if not appropriate))

commercially available, versatile and easy to modify synthetically. n addition, its known ability to enhance proteolytic stability of the attached peptides[9–11], amphiphilicity[12] , conformational properties ensuring correct positioning of the two appended chains[13] and potential to increase cellular uptake and bioavailability[14] make it the perfect dimerizer for this study.

Based on the conclusions of our previous design, we considered the incorporation of a spacer between the peptide and the scaffold in order to provide the final conjugate with enough flexibility to adopt an optimal conformation for specific interaction with the major groove of the target DNA sequence. Moreover, by using of a series of different spacers between the scaffold and the peptide chains we believe to have gained knowledge regarding the factors that can influence DNA binding when synthesizing artificial constructs.

In order to synthesize these constructs in an effective way, we followed a convergent approach involving the CuAAC reaction to conjugate the recognition domains to a functionalized steroid scaffold. The geometric, steric and electronic properties of the 1,2,3-triazole resemble a trans-amide bond, while also affording resistance to enzymatic degradation[15–18], hydrolysis and oxidation, rendering it an attractive heterocyclic moiety to replace physiologically labile linkers in biologically active compounds. In addition, successful replacement of two amino acids in α-helical peptides by a triazole unit has been shown to not significantly influence the secondary peptide structure[15]. To date the triazole linkage has scarcely been used for peptide-steroid conjugation and only to assemble short apolar tripeptides onto bile acid scaffolds[19], despite the increasing interest of these type of conjugates in diverse applications such as HIV inhibitors[20,21] and immunogens for vaccine development[22]. To our knowledge, deprotected, polar peptides have not been assembled on bile acid scaffolds in this manner due to the highly challenging nature of the conjugation. In current work we have optimized the conditions for the CuAAC mediated conjugation of longer, unprotected and functionalized peptides which could be used in the future as a general strategy for the assembly of peptides on a bile acid scaffold.

For this purpose, we have designed and synthesized four different scaffolds for peptide dimerization of the GCN4 basic region. Deoxycholic acid being the common nucleus in the different scaffolds, they differ with respect to the spacers between the peptide and the steroid skeleton, which have different lengths, rigidities and functionalities. Deoxycholic acid was modified at the alcohol positions by attachment of different linkers. The linkers chosen for the study were: pentynoic acid, azido glycine, 4-azidomethyl-benzoic acid and (N-Propynoylamino)-p-toluic acid (PATA). The PATA linker has been specifically developed for bioconjugation purposes as an active alkyne for preparation of peptide-oligonucleotide conjugates via CuAAC[23]. Funtionalization of the steroid nucleus was performed by Steglich esterification, affording final scaffolds (**1**, **2** & **4**). In case of PATA as linker (**3**), the diamino derivative of deoxycholic acid proved necessary for the coupling of the linker, as esterification gave rise to byproducts due to the high reactivity of the alkyne.

The GCN4 basic DNA binding region consists of 23 amino acids that specifically recognize the ATF/CREB-binding site (5'-ATGA C/G TCAT-3'). In order to append the peptides to the bile acid scaffolds, peptides were modified at the C-terminus with unnatural amino acids bearing an alkyne or an azide. Peptides **5** and **6** were synthesized in an automated fashion using Fmoc/tBu SPPS.

The alkyne and azide functionalized GCN4 basic region peptides, **5** and **6** were then attached to the central steroid core, affording four different transcription factor models. Cycloaddition of the peptides to the scaffold via CuAAC was performed in a mixture of DMSO/H_2O, in order to decrease the aggregation of the peptide due to the high polarity of DMSO. Given the hydrophilic nature of the deprotected peptide and the hydrophobic nature of the scaffold, the DMSO/H_2O combination was found to be optimal for the CuAAC reaction. Moreover, it is known that polar solvents prevent aggregation of copper species. As a catalyst, $Cu(CH_3CN)_4PF_6$ gave the best results. A high excess of catalyst was needed presumably due to the complexation of copper with the side chains of the peptide, containing multiple nitrogen containing functionalities including the amine of lysine and the guanidinium group of arginine. Unexpectedly, an excess of scaffold was required for the completion of the reaction, the dipodal construct still being favored over the monopodal one under these conditions. The use of reductants or additives proved not necessary. The reaction was complete after 3 hours at room temperature under argon. Apart from the above mentioned advantages of triazoles, the reaction conditions are compatible with the presence of all unprotected amino acids in the sequence.

Purification of final constructs after completion of the reaction was possible via RP-HPLC, affording compounds **7**, **8**, **9** and **10** in high purity for DNA binding studies.

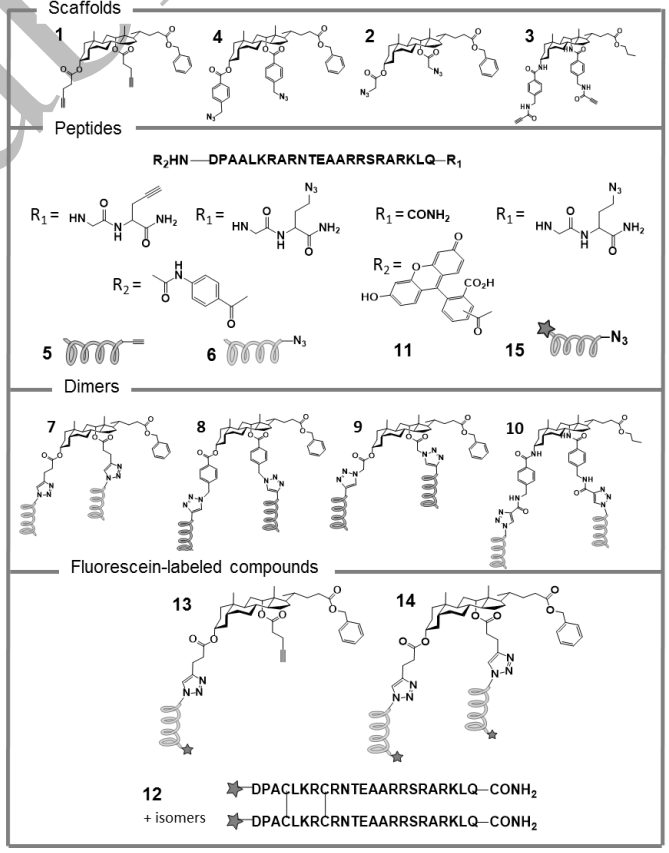


Figure 2. Deoxycholic acid derivatives for the substitution of the dimerization domain with the corresponding peptides.

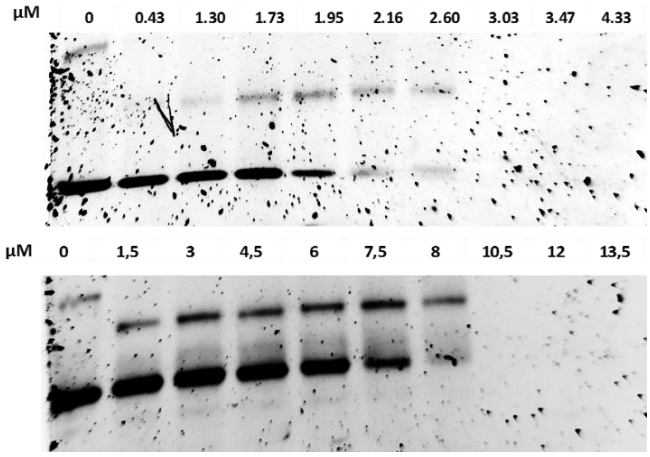


Figure 3. Binding reactions were carried out in the presence of the indicated peptide with <1 nM labeled CRE in a binding mixture (20 pL) containing 20 mM Tris-HCl (pH 7.5). 4 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA, 20 pM poly(d1-dC), and 6% sucrose. The binding mixtures were incubated at 4 "C for 30 min, and an aliquot (5 µL) of each binding mixture was directly loaded onto 8% nondenaturing acrylamide gel (29:1 acrylamide/bis-acrylamide), run in 1/5 TBE buffer (20 mM Tris, 20 mM boric acid, and 0.4 mM EDTA) at 4°C, and analyzed by UV irradiation.

A series of experiments were set up to better understand the cell uptake capacity of the various synthetic constructs. For this purpose the best DNA binding cholic acid based mimic **14** of the GCN4 protein was resynthesized incorporating a fluorescein tag. The properties of this construct were compared to the unmodified GCN4 peptide dimer **11**. The toxicity, quantification of uptake and localisation in the cell of these constructs were studied using RAW264.7 mouse macrophages.

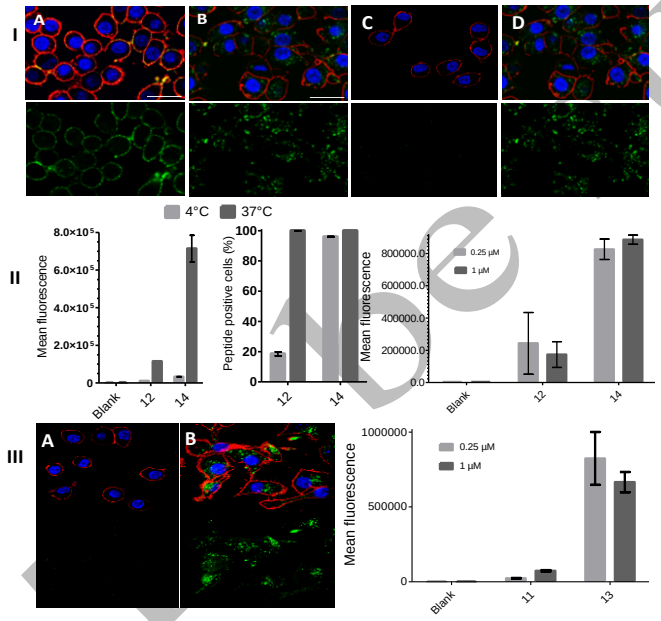


Fig. 4: I. Confocal microscopy images of the uptake of (A) **14** at 4 °C, (B) **14 at** 37°C, (C) **12** at 4 °C, and (D) **12** at 37 °C at 0.25 µM. Upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panel shows fluorescein image. (Scale bar = 20 µm).
II. Mean fluorescence (left) and peptide positive cells (middle) of the synthesized compounds **12** and **14** at 4 °C and 37°C at 0.25 µM. Mean fluorescence of compounds **12** and **14** at 0.25 µM and 1 µM.
III. Confocal microscopy images (left) of the uptake of **11** at (A) 4 °C and (B) 37 °C. Upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). Lower panel shows fluorescein image. (Scale bar = 20 µm). Mean fluorescence (right) of compounds **11** and **13** at 0.25 µM and 1 µM

Although it was shown that all compounds are taken up at 37 °C (figure 4.I), there is clearly an enhanced uptake when deoxycholic acid is used as a scaffold as seen from the mean fluorescence values (figure 4.II). The Monopodal Cholic acid derivative **13**, although non-DNA binding and more hydrophobic, has enhanced uptake like its dimeric counterpart. However, we see a decreased uptake when the concentration is increased to 1 µM. This could be attributed to the denaturant like properties of the cholic acidThe most interesting results were obtained when comparing the localisation of the DNA binding Cholic acid Dimer **14** at 4°C and 37°C (figure 4.IA & 4.IB). At 4°C, where endocytosis is shut down, only insertion in the cell membrane was observed. Accumulation in and around the cell membrane resulted in a high mean fluorescence value and high percentage of peptide positive cells (figure 4.II left and middle). However, at 37°C, whereby both passive and active uptake is possible, the uptake is considerably higher. This indicates that the deoxycholic acid coupled peptides are mainly internalized via active transport at 37°C. However, since we have observed uptake at 4°C for **14** we can conclude that more than one uptake mechanism is involved. The most likely explanation is that the peptides follow an endocytotic pathway whereas the steroid-peptide conjugate performs a direct translocation/transduction type mechanism at 4°C.
As a conclusion, we have illustrated here an optimized CuAAC strategy to conjugate relatively long, deprotected peptides to bile acid scaffolds in a convergent manner. From the four models of the GCN4 bZIP transcription factor presented, the one with a longer and more flexible linker proves to be the best synthetic DNA binder. This can be attributed to the fact that the longer linker allows the construct to grip the major groove of the DNA like a pair of tweezers which is not possible in the case of the less flexible linkers. The binding affinity of **7** is comparable with the earlier described models of GCN4 TFs, however, the synthetic route towards current constructs is considerably simplified. In addition, compound **7** shows a clearly enhanced cell uptake when tested against RAW264.7 mouse macrophages illustrating the particular properties of the steroid nucleus which allow uptake at low concentrations thus affording low cytotoxicity. A steroid in this case is used to increase the uptake of the GCN4 Cell Penetrating Peptide. Thus, the four models of the GCN4 bZIP TF presented here allowed us to detect the parameters that affect dsDNA recognition in synthetic constructs and enable enhanced cell uptake due to the uniqueness of the peptidosteroid conjugates synthesized.

Acknowledgements

Tim Courtin is acknowledged for the determination of the concentration of the final compound by ERECTIC-NMR. Jan Goeman is acknowledged for the LC-MS analysis. Yara Ruiz García and Abhishek Iyer are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. The FWO and BOF are also acknowledged for their financial support.

Keywords: DNA recognition • Steroids • Click chemistry • CPPs • Peptidosteroid conjugation

1] T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell (Cambridge, Mass.)* **1992**, *71*, 1223–1237.
[2] R. V Talanian, C. J. McKnight, P. S. Kim, *Sci. (Washington, D. C., 1883-)* **1990**, *249*, 769–771.

[3] A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo, J. L. Mascarenas, *Angew. Chem., Int. Ed.* **2000**, *39*, 3104–3107.
[4] M. Ueno, A. Murakami, K. Makino, T. Morii, *J. Am. Chem. Soc.* **1993**, *115*, 12575–12576.
[5] J. Mosquera, A. Jimanez-Balsa, V. I. Dodero, M. E. Vazquez, J. L. Mascarenas, *Nat. Commun.* **2013**, *4*, ncomms2825, 8 pp.
[6] L. L. G. Carrette, T. Morii, A. Madder, *European J. Org. Chem.* **2014**, 2883–2891.
[7] B. Cuenoud, A. Schepartz, *Sci. (Washington, D. C., 1883-)* **1993**, *259*, 510–513.
[8] D. Verzele, A. Madder, *Eur. J. Org. Chem.* **2013**, *2013*, 673–687.
[9] C. A. Bode, C. P. Muller, A. Madder, *J. Pept. Sci.* **2007**, *13*, 702–708.
[10] A. Madder, L. Li, M. H. De, N. Farcy, H. D. Van, F. Fant, G. Vanhoenacker, P. Sandra, A. P. Davis, C. P. J. De, *J. Comb. Chem.* **2002**, *4*, 552–562.
[11] C. A. Bode, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller, A. Madder, *Org. Biomol. Chem.* **2009**, *7*, 3391–3399.
[12] W. S. Horne, M. K. Yadav, C. D. Stout, M. R. Ghadiri, *J. Am. Chem. Soc.* **2004**, *126*, 15366–15367.
[13] H. C. Kolb, K. B. Sharpless, *Drug Discov. Today* **2003**, *8*, 1128–1137.
[14] C. W. Tornoe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
[15] I. E. Valverde, A. Bauman, C. A. Kluba, S. Vomstein, M. A. Walter, T. L. Mindt, *Angew Chem Int Ed Engl* **n.d.**
[16] A. Hollmann, P. M. Matos, M. T. Augusto, M. A. R. B. Castanho, N. C. Santos, *PLoS One* **n.d.**, *8*, e60302.
[17] P. Ingallinella, E. Bianchi, N. A. Ladwa, Y.-J. Wang, R. Hrin, M. Veneziano, F. Bonelli, T. J. Ketas, J. P. Moore, M. D. Miller, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5801–5806.
[18] J. P. Tam, *J. Immunol. Methods* **1996**, *196*, 17–32.
[19] M. Wenska, M. Alvira, P. Steunenberg, A. Stenberg, M. Murtola, R. Stroemberg, *Nucleic Acids Res.* **2011**, *39*, 9047–9059.

Cyclodextrin - peptide conjugates for sequence specific DNA binding

Yara Ruiz García[†], Jan Zelenka[‡], Y. Vladimir Pabón Martínez[§], Abhishek Iyer[†], Tomas Kraus[‡], C.I. Edvard Smith[§] and Annemieke Madder*[,†]

†Organic and Biomimetic Chemistry Research Group, Department of Organic and Macromolecular Chemistry, Ghent University, Krijgslaan 281 (S4), B-9000 Ghent (Belgium). Fax: (+32) 9 264 4998. E-mail: annemieke.madder@ugent.be

‡ Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. Flemingovo nam. 2, 166 10 Praha 6. Czech Republic Tel: (+420) 220183372 Fax: (+420) 220183560. E-mail: kraus@uochb.cas.cz

§Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. Tel.: 46-585-83651. Fax: 46-585-83650. E-mail: edvard.smith@ki.

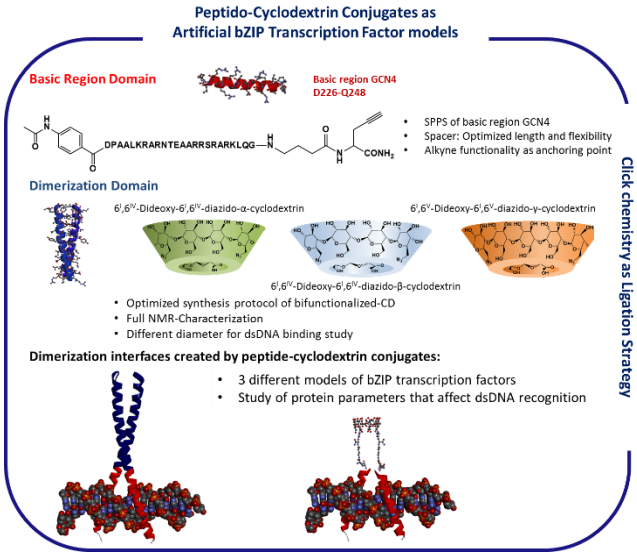
Supporting Information Placeholder

ABSTRACT: Synthetic models of bZIP transcription factors have been developed with the capability of specific DNA recognition. Our design is based on the CuAAC mediated conjugation of basic region Leucine Zipper peptides to different derivatives of α, β and γ-cyclodextrins equipped with azide functionalities. Thorough optimisation of reaction conditions allowed convergent and simultaneous conjugation of two long unprotected cationic peptides to cyclodextrin-bis azide derivatives. The resulting constructs were shown to specifically recognize their cognate DNA sequence with nM affinities. In comparison with previously developed TF models the here described derivatives combine enhanced DNA binding capabilities with an easy and convergent synthetic route.

The study of gene expression regulation is currently of particular interest due to the upcoming development of gene therapy strategies. Indeed, altered expression of particular genes can cause interferences with biological processes in the cell. Therefore, selective up- or down-regulation of specific gene transcription could ultimately result in a therapeutic platform with biomedical applications. During the past decade, biologists and medicinal chemists have been working together in order to target specific genes for this purpose[1–3].

In this context, Transcription Factors (TFs) have been studied in great detail for the modification of gene expression and the study of DNA binding affinity in the cellular environment[4–7]. In particular, bZIP proteins have gained attention due to the simple arrangement of the protein structure consisting of a well-defined dimerization domain and a basic binding region[8]. These proteins are able to interact with DNA in a sequence-specific manner by means of inserting their recognition domains, consisting of defined sequences of amino acids, in the DNA major groove. Simplification of such proteins could ultimately result in peptide-based drugs for alternative disease treatment. Moreover, due to the improvements in peptide manufacturing, peptide drugs can now be produced in a straightforward way through synthetic methods and many techniques have been developed for improved peptide stability.

Therefore, the idea of transferring the protein properties to smaller systems whilst conserving the DNA recognition ability has received considerable attention. The bZIP leucine zipper TF binds dsDNA as a dimer, presented as uninterrupted α-helices that grip the major groove and interact with the DNA through basic residues. The main residues involved in DNA recognition are the amino acids 226-248, located at the N-terminal basic region of the GCN4 protein[8]. The optimal position and spatial arrangement of the dimeric peptides is ensured by the leucine zipper domain. Simplified models of these proteins have been designed based on the substitution of the dimerization domain by a scaffold for the appendage of the basic region peptides in a correct geometry[9]. First proof of this



concept was delivered by Kim et al.[10] using a disulfide bond as connector between the extended basic region of the peptides. Since then, several models have been developed trying to improve upon the pioneer one[11–14].

Figure 1: Design of cyclodextrin based DNA binders.

We here report on cyclodextrin based TF models where the dimerization domain of the GCN4 protein is substituted by a scaffold which allows accurate regulation of the geometry and the mutual distance between the two peptide strands. This has been achieved by the use of α, β and γ-cyclodextrin derivatives in order to scan and evaluate the importance of the distance between the two anchoring points for the dimerizing peptides (figure 1). Due to the rigid structure of the cyclodextrins, attachment of the basic region peptide of the GCN4 protein onto opposite positions of the primary rim of the cyclodextrins indeed allows control of the dimerization distance. Based on the diameters of the primary rim of α, β and γ–cyclodextrins, which are 5.7, 7.8 and 9.5 Å respectively[15], three different peptide-cyclodextrin conjugates have been synthesized, keeping in mind that the width of the major groove is 11.2 Å. For this purpose, we have synthesized $6^{I},6^{IV}$-diazido-α-cyclodextrin, $6^{I},6^{IV}$-diazido-β-cyclodextrin and $6^{I},6^{V}$ - diazido-γ-cyclodextrin, (Figure 2), in order to conjugate peptides which are functionalized with an alkyne at the C-terminus via CuAAC.

In view of the specific geometry, steric and electronic properties, a 1, 2, 3-triazole can be regarded as a trans-amide bond mimic[16]. Moreover, this linkage is stable under physiological conditions, thus representing a perfect heterocyclic moiety to replace unstable linkers[17–19]. In addition, successful replacement of dipeptide sequences in α-helical peptides by triazole units has been shown to not significantly influence the secondary peptide structure[16].

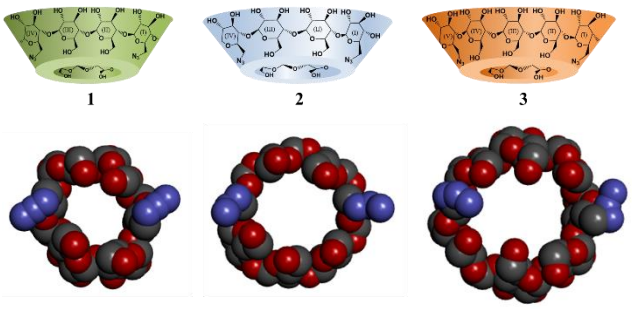
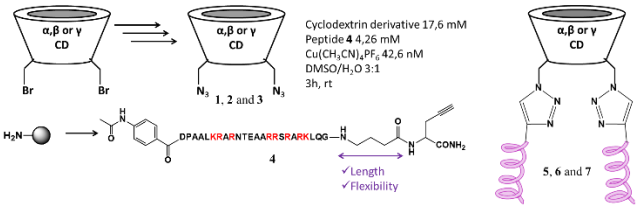


Figure 2. Molecular visualization of the cyclodextrin derivatives synthesized as dimerization domains. From left to right: 6I,6IV-diazido-α-cyclodextrin **1**, 6I,6IV-diazido-β-cyclodextrin **2** and 6I,6V-diazido-γ-cyclodextrin **3**.

Peptide-cyclodextrin conjugates have been previously developed for a broad range of applications, such as drug release systems[20,21], enantioselective ester hydrolysis catalysis[22], and for the creation of new types of chemosensors[23], enzyme mimics[24] and self-assembling materials[25,26]. Although CD conjugation to peptides has been intensively studied, we here for the first time use the cyclodextrin moiety as artificial dimerizing unit for mimicking protein-DNA interactions.

The syntheses of the diazido cyclodextrin derivatives **1** and **2** via $6^{I},6^{IV}$-dichloro-α-cyclodextrin and $6^{I},6^{IV}$-diiodo-β-cyclodextrin[27] precursors, respectively, were described earlier [28]. Nevertheless, we used $6^{I},6^{IV}$-dibromo-α-cyclodextrin[29], $6^{I},6^{IV}$-dibromo-β-cyclodextrin[30] and $6^{I},6^{V}$-dibromo-γ-cyclodextrin[31] and converted



them in one step to compounds **1-3** (Figure 3) by reaction with three-fold excess of sodium azide in DMF at 50°C. Pure compounds **1-3** were isolated from the reaction mixtures by reversed-phase column chromatography with yields of 90%, 83% and 77% for the α, β and γ-cyclodextrin derivatives respectively.

Figure 3. Synthesis scheme for the conjugation of the different cyclodextrins 1, 2 and 3 and peptide 4 to obtain final compound 5, 6 and 7. Residues that can chelate copper are marked in red.

As we have previously illustrated, peptides attached directly to a dimerization unit are unable to adopt an adequate position for major groove binding[32]. Thus, a spacer is needed between the dimerization scaffold and the basic region GCN4 peptide to afford some flexibility required for binding. In addition to that, a C-terminal propargylglycine residue was introduced for conjugation of the deprotected peptide to the azide containing cyclodextrin scaffolds. The GCN4 basic region consists of the following 23 amino acids: DPAALKRARNTEAARRSRARKLQ which specifically recognize the ATF/CREB-binding site (5'-ATGA C/G TCAT-3'). The monomeric GCN4 sequence was synthesized using standard Fmoc SPPS on a Rink-amide ChemMatrix resin. After cleavage and deprotection, the peptide could be used for the next steps without intermediate purification (Figure 4).

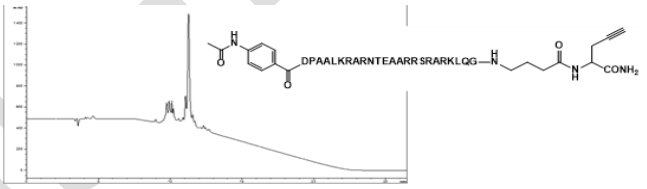
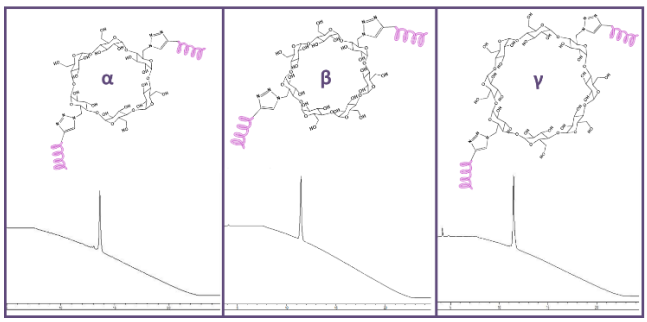


Figure 4. RP-HPLC Chromatogram at 214 nm of crude peptide 4. (0-100% ACN in 15 min on C4, 300 Å).

The basic region GCN4 peptide functionalized with an alkyne at the C-terminus was then anchored to opposite positions at the primary rim of α, β and γ-cyclodextrin derivatives (Figure 3). The Cu-AAC is a broadly developed strategy for the bioconjugation of peptides to different building blocks due to its mildness, compatibility with a large variety of functional groups and the possibility of performing the reaction in wide range of solvents and buffers. In order to prevent aggregation of the peptide during the reaction, high polarity media were needed. In addition, polar solvents prevent copper species from aggregation. Therefore, we chose DMSO/H_2O as a solvent mixture. It is known that sodium ascorbate can cause covalent modifications of lysines and arginines, and can form reactive oxygen species [33,34]. Therefore, $Cu(CH_3CN)_4PF_6$ was chosen as catalyst after carefully studying different reaction conditions. It is to be noted that a high excess of catalyst is needed which can probably be ascribed to the chelation of copper by the nitrogen containing amino acid side chains, which are intensely represented within this particular DNA binding peptide sequence. The reaction was



stopped after 3 hours at room temperature under argon. Final purification of the reaction mixture was achieved via RP-HPLC and all conjugates could be obtained in high purity (Figure 5).

Figure 5. Structure of final compounds 5 (α), 6 (β) and 7 (γ) and RP-HPLC Chromatograms of purified compounds (0-100% ACN in 15 min on C4, 300 Å).

After successful synthesis of the constructs, the DNA binding capabilities of the systems were investigated. Quantitative binding analysis was performed by EMSA using double stranded ^{32}P-labeled CRE DNA (CRE: 5'-ATGACGTCAT-3'), which is the natural palindromic binding site of GCN4 protein. This technique allows the sensitive determination of the dissociation constant (Kd) (figure 6).

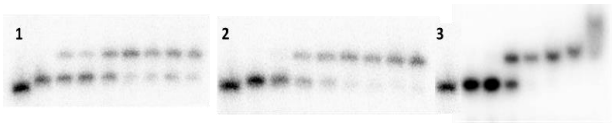


Figure 6. EMSA titration of the dipodal peptidocyclodextrin conjugates to the 5'-labeled ^{32}P-CRE sequence (CRE: 5'-ATGACGTCAT-3'): Concentrations: 5 nM dsDNA; First lane in all the gels: pyrimidine strand. Concentration of peptide in the lanes 2-9 for gel 1 and 2 (5 and 6): 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 µM. Concentration of peptide in the lanes 2-8 for gel 3 (7): 0, 0.06, 0.125, 0.25, 0.5, 1, 2 µM. A Fuji FLA3000 phosphorimager was used for gel analysis and Multi Gauge V 3.0 software (Fujifilm) for quantification of the electrophoretic band intensities.

The results show all conjugates bind to the target DNA sequence as evident from the appearance of an up-shifted band. In contrast no binding was observed when incubating the conjugates with random DNA sequences (see SI). Previous models developed by Morii and Mascareñas possess the same length of the basic region peptides (D226-Q248) and are therefore comparable with our systems. The group of Morii developed a system in which the dimerization unit was a complex of cyclodextrin and adamantane. Both counterparts were anchored to a GCN4 basic region peptide and bind only when dimerized. EMSA studies were performed with < 1 nM radiolabelled CRE, obtaining a Kd < 150 nM[11]. More recent models designed by the group of Mascareñas employed a diazobenzene moiety for dimerization, with a Kd < 5 nM using < 1 nM radiolabelled CRE. They also reported the affinity of a disulfide-bridged GCN4 basic region, with a Kd of <150 nM approximately for a DNA concentration of 50 nM[35]. The latest model consisted on a dual mimic in which dimerization arrangement and DNA selectivity was controlled by selected external parameters[36]. The basic region peptides were functionalized with a cysteine and a terpyridine moiety attached to the residues at the N- and C- termini of the peptide chain to achieve selective dimerization at both sites. Depending on the conditions, the disulfide-based or the metal-terpyridine complex-based dimers were preferred, and different DNA sequences could be targeted. By dimerizing the peptide in the presence of Ni^{2+}, the construct was bound to the CRE sequence with a Kd of 299±26 nM, using a DNA concentration of 100 nM, ~100 pM labeled with ^{32}P. For these 3 last models, it should be noted that the nonionic detergent (NP-40) and BSA were used during EMSA, known to decrease aggregation of peptides and proteins and therefore favor the interaction with the DNA[37,38]. In our case, the obtained Kd values for compounds **5**, **6** and **7** were 50 ± 20, 30 ± 20 and 100 ± 60 nM respectively. A rough comparison thus indicates that our new constructs, apart from being synthesized in a straightforward and convergent manner, display comparable binding properties to those described in literature. The obtained values are further within the same order of magnitude as those calculated for the binding of bacterially expressed GCN4 and synthetic versions thereof. For instance, a dimer comprising of a 56mer GCN4 basic region (residues 226-281) also binds CRE in the nanomolar range (Kd ~12 nM)[39–41]. No data for the dissociation constant of the natural protein have been reported so far.

In order to better explain the results obtained from the binding pattern on the gels and the determination of the Kd values, we performed a molecular visualization of the dimerization interface of the bZIP GCN4 TF obtained by discarding the leucine zipper domain from the crystal structure of the natural protein (figure 7, D). The distance between the C-termini of the basic regions is shown in figure 7 at different perspectives of the protein-DNA interface, and is found to be 15.052 Å. The Kd values showed that derivative **6** was the one with the best binding capability to the CRE sequence. Therefore, we consider that beta CD is the dimerizer which allows optimal anchoring of the peptides on the major groove of the DNA. In the protein the C-termini are placed at a distance of approximately 15 Å. In case of our constructs, although the diameter of the primary rim of the cyclodextrins is known (figure 7, left), due to the flexibility of the linker, the exact distance between the C-termini of the peptides on **5**, **6** and **7** could not be predicted accurately. However, we observed that increasing or decreasing the distance between the attachment points of the peptides by the use of different CD scaffolds results in a deviation of the optimum geometry of the system. This is reflected in terms of decreased binding affinity and lower Kds. It should also be noted that these results are specific for the basic region GCN4 peptide sequence and the given spacers. In case these parameters are varied, different conclusions are expected.

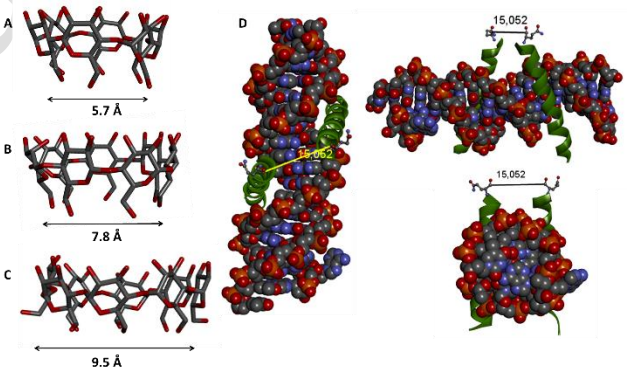


Figure 7. 3D structure of α, β and γ-cyclodextrin and primary rim diameter distance (A, B and C respectively) and crystal structure of the GCN4 basic region peptides appended to the major groove of the DNA from different perspectives. Distance between C-termini indicated in angstroms (PDB: 1YSA) (D).

In conclusion, three peptide-cyclodextrin conjugates for sequence-selective DNA recognition were obtained. This was achieved by the use of α, β and γ-cyclodextrin diazido derivatives as scaffolds for the appendage of the peptides by CuAAC. We have successfully synthesized and fully characterized $6^{I},6^{IV}$-diazido-α-cyclodextrin (**1**), $6^{I},6^{IV}$-diazido-β-cyclodextrin (**2**) and 6^{I}, 6^{V}-diazido-γ-cyclodextrin (**3**). Though examples exist of CuAAC reactions with CD derivatives [25,42], to the best of our knowledge long, deprotected

peptides have so far not been conjugated to cyclodextrins via Cu-AAC. We here present optimized conditions for the anchoring of such long peptides to cyclodextrin units via click chemistry.

Our results indicate the usefulness of an optimized dimerization configuration between both peptides in artificial TF models. Indeed, it was shown that the distance between the anchoring points has a notable influence on DNA binding. Successful models can be obtained by trying to approach the exact features of the protein at the interface between the dimerization and the basic region domains of the bZIP TF to achieve DNA binding comparable to that of the natural TF.

ASSOCIATED CONTENT

Supporting Information

Reaction conditions, RP-HPLC traces, ESI-MS, NMR spectra and EMSA studies. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: annemieke.madder@ugent.be.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

Tim Courtin is acknowledged for the determination of the concentration of the final compounds by ERECTIC-NMR. Jan Goeman is acknowledged for the LC-MS analysis. Yara Ruiz García and Abhishek Iyer are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. Support from Ministry of Education, Youth and Sport of the Czech Republic to TK (LD12019) is greatly appreciated.

REFERENCES

(1) Goverdhana, S.; Puntel, M.; Xiong, W.; Zirger, J. M.; Barcia, C.; Curtin, J. F.; Soffer, E. B.; Mondkar, S.; King, G. D.; Hu, J.; Sciascia, S. A.; Candolfi, M.; Greengold, D. S.; Lowenstein, P. R.; Castro, M. G. *Mol. Ther.* **2005**, *12*, 189.

(2) Klug, A. *Annu. Rev. Biochem.* **2010**, *79*, 213.

(3) Li, X.; Chen, S.; Sun, T.; Xu, Y.; Chen, Y.; Liu, Y.; Xiang, R.; Li, N. *Clin. Lab.* **2014**, *60*, 909.

(4) Ma, P. C.; Rould, M. A.; Weintraub, H.; Pabo, C. O. *Cell* **1994**, *77*, 451.

(5) Glover, J. N.; Harrison, S. C. *Nature* **1995**, *373*, 257.

(6) Shimizu, T.; Toumoto, A.; Ihara, K.; Shimizu, M.; Kyogoku, Y.; Ogawa, N.; Oshima, Y.; Hakoshima, T. *EMBO J.* **1997**, *16*, 4689.

(7) Darnell Jr., J. E. *Nat. Rev. Cancer* **2002**, *2*, 740.

(8) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell (Cambridge, Mass.)* **1992**, *71*, 1223.

(9) Pazos, E.; Mosquera, J.; Vázquez, M. E.; Mascareñas, J. L. *Chembiochem* **2011**, *12*, 1958.

(10) Talanian, R. V; McKnight, C. J.; Kim, P. S. *Sci. (Washington, D. C., 1883-)* **1990**, *249*, 769.

(11) Ueno, M.; Murakami, A.; Makino, K.; Morii, T. *J. Am. Chem. Soc.* **1993**, *115*, 12575.

(12) Cuenoud, B.; Schepartz, A. *Sci. (Washington, D. C., 1883-)* **1993**, *259*, 510.

(13) Caamano, A. M.; Vazquez, M. E.; Martinez-Costas, J.; Castedo, L.; Mascarenas, J. L. *Angew. Chem., Int. Ed.* **2000**, *39*, 3104.

(14) Carrette, L. L. G.; Morii, T.; Madder, A. *European J. Org. Chem.* **2014**, *2014*, 2883.

(15) Brewster, M. E.; Loftsson, T. *Adv. Drug Deliv. Rev.* **2007**, *59*, 645.

(16) Horne, W. S.; Yadav, M. K.; Stout, C. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2004**, *126*, 15366.

(17) Kolb, H. C.; Sharpless, K. B. *Drug Discov. Today* **2003**, *8*, 1128.

(18) Tornoe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(19) Valverde, I. E.; Bauman, A.; Kluba, C. A.; Vomstein, S.; Walter, M. A.; Mindt, T. L. *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 8957.

(20) Schaschke, N.; Assfalg-Machleidt, I.; Machleidt, W.; Laßleben, T.; Sommerhoff, C. P.; Moroder, L. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 677.

(21) Decroocq, C.; Joosten, A.; Sergent, R.; Mena Barragán, T.; Ortiz Mellet, C.; Compain, P. *Chembiochem* **2013**, *14*, 2038.

(22) Tsutsumi, H.; Ikeda, H.; Mihara, H.; Ueno, A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 723.

(23) Hossain, M. A.; Hamasaki, K.; Takahashi, K.; Mihara, H.; Ueno, A. *J. Am. Chem. Soc.* **2001**, *123*, 7435.

(24) Tsutsumi, H.; Hamasaki, K.; Mihara, H.; Ueno, A. *J. Chem. Soc. Perkin Trans. 2* **2000**, 1813.

(25) Lin, Y.-C.; Wang, P.-I.; Kuo, S.-W. *Soft Matter* **2012**, *8*, 9676.

(26) Beck, J. N.; Singh, A.; Rothenberg, A. R.; Elisseeff, J. H.; Ewald, A. J. *Biomaterials* **2013**, *34*, 9486.

(27) Reineke, T. M.; Davis, M. E. *Bioconjug. Chem.* **2002**, *14*, 247.

(28) Guitet, M.; Marcelo, F.; de Beaumais, S. A.; Zhang, Y.; Jiménez-Barbero, J.; Tilloy, S.; Monflier, E.; Ménand, M.; Sollogoub, M. *European J. Org. Chem.* **2013**, *2013*, 3691.

(29) Kumprecht, L.; Budesínský, M.; Vondrásek, J.; Vymetal, J.; Cerný, J.; Císarová, I.; Brynda, J.; Herzig, V.; Koutník, P.; Závada, J.; Kraus, T. *J. Org. Chem.* **2009**, *74*, 1082.

(30) Grishina, A.; Stanchev, S.; Kumprecht, L.; Buděšínský, M.; Pojarová, M.; Dušek, M.; Rumlová, M.; Křížová, I.; Rulíšek, L.; Kraus, T. *Chemistry* **2012**, *18*, 12292.

(31) Volkov, S.; Kumprecht, L.; Buděšínský, M.; Dušek, M.; Kraus, T. *To be Submitt.*

(32) Verzele, D.; Madder, A. *Eur. J. Org. Chem.* **2013**, *2013*, n/a.

(33) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem., Int. Ed.* **2009**, *48*, 9879.

(34) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. *Bioconjugate Chem.* **2010**, *21*, 1912.

(35) Jiménez-Balsa, A.; Pazos, E.; Martínez-Albardonedo, B.; Mascareñas, J. L.; Vázquez, M. E. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8825.

(36) Mosquera, J.; Jimanez-Balsa, A.; Dodero, V. I.; Vazquez, M. E.; Mascarenas, J. L. *Nat. Commun.* **2013**, *4*, ncomms2825.

(37) Hellman, L. M.; Fried, M. G. *Nat. Protoc.* **2007**, *2*, 1849.

(38) *Cell Biology (Page 330)*; Third.; Elsevier: Oxford (UK), 2006; p. 584.

(39) Hollenbeck, J. J.; Oakley, M. G. *Biochemistry* **2000**, *39*, 6380.

(40) Bird, G. H.; Lajmi, A. R.; Shin, J. A. *Biopolymers* **2002**, *65*, 10.

(41) Chan, I.-S.; Fedorova, A. V; Shin, J. a. *Biochemistry* **2007**, *46*, 1663.

(42) Faugeras, P.-A.; Boëns, B.; Elchinger, P.-H.; Brouillette, F.; Montplaisir, D.; Zerrouki, R.; Lucas, R. *European J. Org. Chem.* **2012**, *2012*, 4087.

ChemComm

RSCPublishing

COMMUNICATION

Specific Recognition of a Non-palindromic dsDNA Sequence by a Heterodimeric Transcription Factor Mimic

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Yara Ruiz García,[a] Vladimir Pabón Martínez,[b] Edvard Smith[b] and Annemieke Madder*[a]

We have designed and synthesized the first mimic of the c-Myc/Max-bHLH-ZIP transcription factor able to selectively recognize the dsDNA sequence 5'-CACGTG-3' and complement on the major groove of DNA. Our design is based on our previouspeptide-steroid conjugates which are proven to be effective as DNA binders and afford enhanced bioavailability and cell-permeability.

In the last decades, the development in the areas of biotechnology and chemical biology has opened up new prospects in anti-cancer research. Among other promising approaches that have recently gained attention, the inhibition of oncogenes and/or the reactivation of tumor suppressor genes by transcription factors (TFs) are of particular interest. These TFs are responsible for maintenance and progress of tumor growth[1,2] and can be regarded as desirable targets in gene therapeutic strategies.

One of the major families of transcription factors related to human cancer and consequently targeted for cancer therapy is the Myc/Max/Mad basic (Helix-Loop-Helix) zipper transcription factor (bHLH-ZIP TF) network. These oncoproteins are heterodimers composed of proteins from the Myc, Max and Mad families[3–5]. The c-Myc-Max heterodimer binds DNA in a sequence-specific manner and its function is the transcription from promoters containing the E-box sequence 5'-CACGTG-3'[6]. The disregulation of the c-Myc gene causes its overexpression that is related to the progression of diseases such as Burkitt's lymphoma, cancer in infancy as neuroblastomas and lung, breast and colon carcinomas[7,8].

The crystal structure of cMyc-Max bHLH-ZIP transcription factor as an oncoprotein has been studied in detail. Eukaryotic bZIP and bHLH-ZIP TF motifs have one of the simplest protein structures capable of sequence-specific DNA recognition on the major groove. The structural motif of bHLH-ZIP is similar to the bZIP one. Both are characterized by a dimerization domain, which is a prerequisite for DNA binding and a DNA recognition domain known as the basic region that contains the α-helical peptides responsible for the specific DNA binding. However, the main difference is the interface between the two domains. In the case of bZIP proteins the α helices are continuous, while in the case of bHLH-ZIP proteins they are interrupted by a loop making the design and synthesis of the mimics of the bHLH-ZIP protein considerably more complicated (figure X). The Helix-Loop-Helix motif does not contribute significantly to DNA binding or to dimerization and is known to tolerate variations in size[9,10].

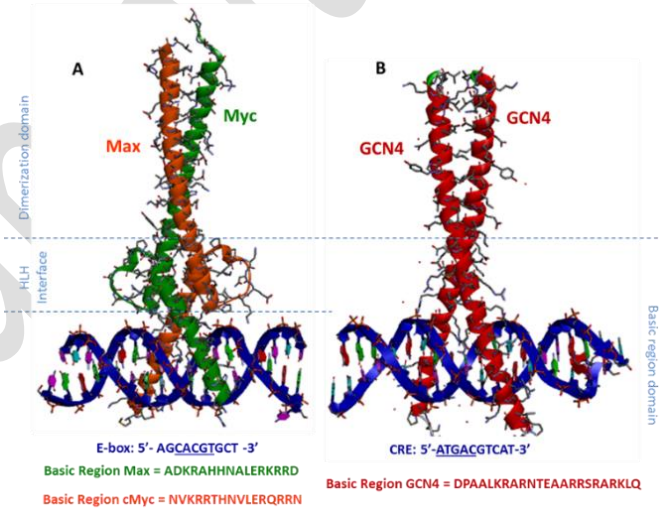
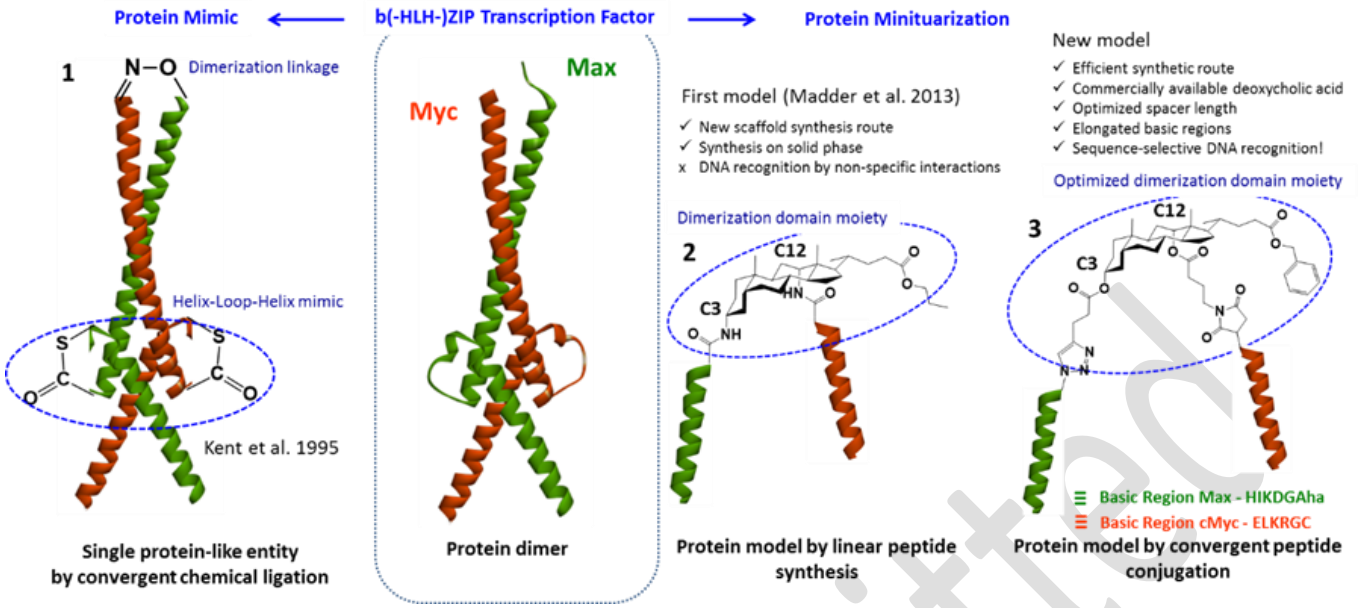


Figure X. Experimentally-determined structures of cMyc/Max bHLH-ZIP (left) and GCN4 bZIP (right) Transcription Factors.

Due to the relevant role of this oncoprotein in cancer development, Kent et al[11] developed the first approach directed towards the total synthesis of a non-covalently-linked heterodimeric bHLH-ZIP protein mimic by convergent chemical ligation of unprotected peptide segments. However, progress in the synthesis of the simplified heterodimeric bHLH-ZIP TF models has only been investigated by the OBCR group, by the substitution of the dimerization domain via a steroid-based moiety[12] (Figure X). On the other hand, several artificial models of bZIP transcription factors have been developed by various



groups[13–18], by substituting the dimerization domain for a synthetically accessible linker to which the basic regions were attached.

Figure X. Illustration of the first mimic of bHLH-ZIP TF by Kent (left) and our cMyc/Max bHLH-ZIP TF models (right).

Steroidal building blocks have been extensively studied in drug design due to their amphiphilic nature, bioavailability and tendency to enhance peptide biostability[19–22]. In addition, due to their rigid core, the right dimensions for major groove binding increase the helicity of the appended peptides and they can substitute the leucine zipper domain as an anchoring point for the basic region of bHLH ZIP TF. Consequently, peptidosteroid tweezers are perfect alternatives to current chemotherapeutics, leading to a new type of sequence-specific DNA binder with less toxic side effects by preventing the interaction of a TF with its promoter.

In our group, the previously synthesized artificial peptidosteroid-based mimic of this oncoprotein lead to unsuccessful DNA binding due to the lack of an α-helix conformation in the peptides. This lack of α-helicity resulted in the displacement of the side chains of the peptide with respect to the DNA, disrupting the specific interactions involved in the peptide-DNA complex [12]. As reported in the article, the direct conjugation of the basic region to the scaffold impedes the correct dimerization of the peptides due to the rigidity and sterical encumbrance imposed by the nucleus of the steroid scaffold. This was illustrated by the high resolution of the protein structure by Burley et al[3](PDB-ID: 1NKP). As a conclusion, we proposed the need for a spacer between the dimerization moiety in addition to having an elongated DNA recognition domain in our artificial DNA binder. This would then make the mimic a better structural fit in comparison to the natural protein. We considered the elongation of the basic region peptides at the C-terminus corresponding to at least four amino acids or approximately one helical turn which should be added to the 17-residue c-Myc/Max peptides. A linker was incorporated on the final peptide-steroid conjugate in order to provide the peptides with the autonomy needed to be inserted to the major groove of the DNA in the optimal configuration. Glycine was coupled as second residue to increase the space between the scaffold and the peptide as reported in previous models of bZIP TF. Moreover, a convergent approach was applied for the appendage of the peptides onto the scaffold, as a more efficient way to synthesize supramolecular peptide conjugates. Based on our previous results, we hereby report the synthesis of an optimized model of bHLH ZIP transcription factor that binds specifically to the non-palindromic target DNA sequence.

In order to dimerize the peptides c-Myc and Max onto the scaffold, an orthogonally-functionalized deoxycholic acid scaffold derivative was synthesized to allow the convergent incorporation of the basic region peptides of c-Myc and Max oncoproteins. The design and synthesis of a deoxycholic acid derivative scaffold for peptide heterodimerization of the Myc/Max basic region was possible due to the concave structure of the steroid nucleus. This allowed the two alcohol functionalities to be selectively esterified to allowing convergent conjugation of two different peptide sequences[24]. The selective esterification at C3 with pentynoic acid was carried out using catalytic amount of DMAP and EDC·HCl as coupling reagent. Subsequent functionalization of the alcohol at C12 with Boc protected gamma-aminobutyric acid was performed with EDC and excess of DMAP to lead the reaction to completion. Finally, Boc deprotection and conversion of the obtained amine to maleimide resulted in the hetero-functionalized scaffold. As reported previously, an aliphatic linker is needed for DNA binding in the area of transcription factor mimics[23]. The spacers between the peptides and the scaffold should provide sufficient length and flexibility to the mimic to accommodate the peptides into the DNA major groove in the right conformation. Therefore, we chose 4-maleimido-butyric acid and pentynoic acid as linkers to allow the flexibility needed for DNA recognition at the major groove.

The Click and cysteine maleimide conjugation strategies were chosen for peptide conjugation due to their generality and orthogonality. For this purpose, the c-Myc and Max peptides sequences were modified at the C-terminus with a cysteine and azido-homoalanine (Aha) respectively. The Myc/Max basic regions consist of a sequence of 16 aminoacids that specifically recognize the E-box DNA sequence (5'- AGCACGTGCT -3'). The monomeric Myc and Max sequences, were synthesized using Fmoc-/tBu-SPPS on Rink-amide ChemMatrix resin. The Myc/Max basic regions, functionalized with a cysteine and an

azide, were conjugated to the central steroid scaffold resulting in the heterodimeric TF mimic. First, the conjugation at the C12 position was performed via thiol-maleimide conjugation through 1, 4 Michael addition to a cysteine-contained peptide resulting in a succinimidyl thioether moiety. The second conjugation was performed via CuAAC at C3 position of the alkyne-functionalized scaffold forming a triazole. The click conjugation was compatible with the presence of the peptide chain at C12.
The conjugation strategy is based on our previous results for the conjugation of peptides to steroid scaffolds in which we explored the CuAAC for the attachment of long deprotected peptides to hydrophobic cores. Our successful results made click chemistry the best choice for functionalization of the challenged C3 position due to the steric constraint imposed by the deprotected peptide at C12 position as a result of hydrogen-bond and electrostatic interactions with the free peptide in solution and the masking of the exposure of the alkyne for the conjugation to the azide in the peptide. Moreover, the results show that our CuAAC protocol is compatible with different peptide sequences. Both, the succinimidyl thioether and triazole moieties are generally accepted as stable linkages under physiological conditions.

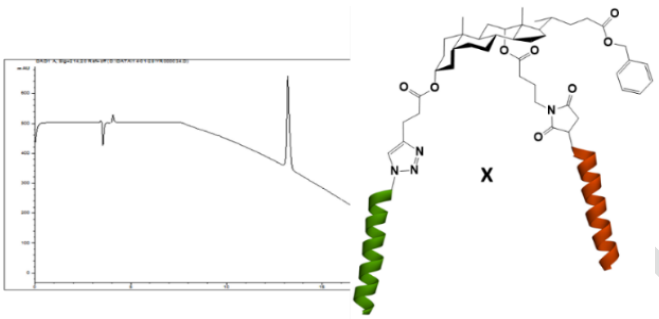


Figure X. RP-HPLC Chromatogram of purified compound X.

DNA RECOGNITION

The heterodimeric c-Myc/Max peptidosteroid conjugate X after purification was subjected to EMSA study with ^{32}P-radiolabelled DNA, in order to test the recognition capability to the E-box sequence (5' –CA CGT G – 3'). Therefore, the symmetric sequence (5' – CTA CTA GCA CGT GCT AGT AG – 3') was annealed to form the DNA-duplex subjected to complex formation with the peptides. The results of the experiment clearly reflect the formation of dsDNA-dimeric peptide construct due to the presence of an upper band with lower mobility. The presence of two bands at the dsDNA lane is consequence of the incomplete annealing of both DNA sequences. The lowest band corresponds to the ssDNA while the upshifted band is the dsDNA (figure X-gel from Vladimir). The dissociation constant of the complex formation is … nM.

To conclude, the potential of a reduced-size bHLH-ZIP cMyc/Max TF to recognize its dsDNA binding sequence was tested. The results show that an optimized and miniaturized version of the protein contains the necessary features to form a complex with dsDNA in a sequence-specific manner (figure X). In addition, we show the possibility of selective functionalization of steroid-based scaffolds to synthesize heterodimeric peptide-based conjugates, broadening the application of current methodology to other families of TF, such as bZIP Fos/Jun TF. In addition to the above, our design overcomes the difficulties of the decoration of building blocks by the well-known SPPS, since it is based solely on conjugation strategies in the solution phase. This can be considered the first model of an oncoprotein accessible by synthetic methods towards cancer therapy development.

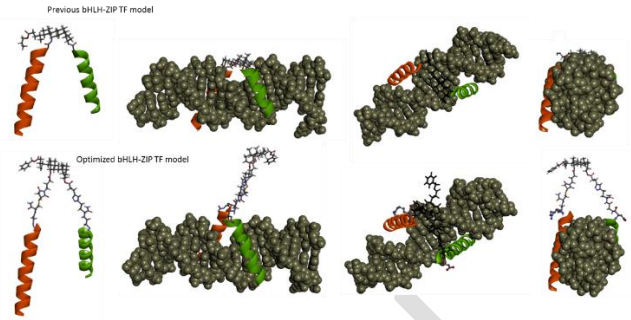


Figure X. Visualization models of our previous peptide-steroid conjugate and the optimized one reported in this article with its binding sequence from different perspectives.

1. J. E. Darnell Jr., *Nat. Rev. Cancer*, 2002, **2**, 740–749.
2. T. A. Libermann and L. F. Zerbini, *Curr. Gene Ther.*, 2006, **6**, 17–33.
3. S. K. Nair and S. K. Burley, *Cell (Cambridge, MA, U. S.)*, 2003, **112**, 193–205.
4. V. H. Cowling and M. D. Cole, *Semin. Cancer Biol.*, 2006, **16**, 242–252.
5. N. Kuramoto, K. Ogita, and Y. Yoneda, *Jpn. J. Pharmacol.*, 1999, **80**, 103–109.
6. E. M. Blackwood and R. N. Eisenman, *Sci. (Washington, D. C., 1883-)*, 1991, **251**, 1211–1217.
7. M. Vita and M. Henriksson, *Semin. Cancer Biol.*, 2006, **16**, 318–330.
8. C. V Dang, *Cell (Cambridge, MA, U. S.)*, 2012, **149**, 22–35.
9. L. E. Canne, A. R. Ferre- D'Amare, S. K. Burley, and S. B. H. Kent, *J. Am. Chem. Soc.*, 1995, **117**, 2998–3007.
10. A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff, and S. K. Burley, *Nat.*, 1993, **363**, 38–45.
11. L. E. Canne, A. R. Ferre- D'Amare, S. K. Burley, and S. B. H. Kent, *J. Am. Chem. Soc.*, 1995, **117**, 2998–3007.
12. D. Verzele and A. Madder, *Eur. J. Org. Chem.*, 2013, **2013**, 673–687.
13. R. V Talanian, C. J. McKnight, and P. S. Kim, *Sci. (Washington, D. C., 1883-)*, 1990, **249**, 769–771.
14. B. Cuenoud and A. Schepartz, *Sci. (Washington, D. C., 1883-)*, 1993, **259**, 510–513.
15. M. Ueno, A. Murakami, K. Makino, and T. Morii, *J. Am. Chem. Soc.*, 1993, **115**, 12575–12576.
16. A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo, and J. L. Mascarenas, *Angew. Chem., Int. Ed.*, 2000, **39**, 3104–3107.
17. J. Mosquera, A. Jimanez-Balsa, V. I. Dodero, M. E. Vazquez, and J. L. Mascarenas, *Nat. Commun.*, 2013, **4**, ncomms2825, 8 pp.
18. L. L. G. Carrette, T. Morii, and A. Madder, *European J. Org. Chem.*, 2014, **2014**, 2883–2891.
19. T. Vorherr, *PharmaChem*, 2008, **7**, 13–16.
20. X.-Z. Lai, Y. Feng, J. Pollard, J. N. Chin, M. J. Rybak, R. Bucki, R. F. Epand, R. M. Epand, and P. B. Savage, *Acc. Chem. Res.*, 2008, **41**, 1233–1240.
21. E. Virtanen and E. Kolehmainen, *Eur. J. Org. Chem.*, 2004, 3385–3399.
22. C. A. Bode, C. P. Muller, and A. Madder, *J. Pept. Sci.*, 2007, **13**, 702–708.

23. Y. Ruiz García, A. Iyer, V. Pabón Martínez, D. Van Lysebetten, B. de Geest, E. Smith, and A. Madder, *Angew. Chem., Int. Ed.*, 2014.
24. M. G. Simpson, M. Pittelkow, S. P. Watson, and J. K. M. Sanders, *Org. Biomol. Chem.*, 2010, **8**, 1181–1187.

Conclusions

The conclusions section should come at the end of the article.

Notes and references

[a] Organic and Biomimetic Chemistry Research Group, Krijgslaan 281, S4, B-9000. Gent, Belgium. E-mail: annemieke.madder@ugent.be; Fax:+32-9-264 49 98.
[b]Clinical Research Center, Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86, Stockholm, Sweden. Tel.: 46-585-83651. Fax: 46-585-83650. E-mail: edvard.smith@ki.se

†Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/c000000x/

1 Citations here in the format A. Name, B. Name and C. Name, *Journal Title*, 2000, **35**, 3523; A. Name, B. Name and C. Name, *Journal Title*, 2000, **35**, 3523.

DOI: 10.1002/anie.200((will be filled in by the editorial staff))

Untapped Opportunities of Resin-to-Resin Transfer Reactions (RRTR) for the Convergent Assembly of Multivalent Peptide Conjugates**

*Yara Ruiz García, Dieter Verzele and Annemieke Madder**

Over the past decade, bioconjugation strategies has been increasingly developed in the fields of bioorganic chemistry and chemical biology, along with the rise of the bioorthogonal chemistry concept.[1,2] Developments include tandem, one-pot or solid-phase chemical ligations (SPCL),[3,4] and methodologies are currently as sophisticated as, for example, dual kinetically controlled ligations (KCL) for the assembly of peptides.[5] The predominance of Kent et al.'s native chemical ligation (NCL) method[6,7] confirms the desire for native amide bonds, given their omnipresence in life and nature, with recent focus towards conceptual advances in amidation strategies.[8]

Despite the merits of these methods, several issues persist and should not be trivialized. Difficulties often arise during handling of the precursor peptide segments, as agreed by important peers in this area[9] and encountered in our daily practice, despite ongoing efforts reported in literature to overcome these issues.[10] With the need for specific reactive handles, synthesis of multivalent designs by ligation methods furthermore requires significant strategic scheduling to obtain a desired assembly. Therefore, we felt a need to return to simplicity from a practical perspective focussing on the solid-phase resin as keystone of peptide synthesis. We here wish to describe so-called resin-to-resin transfer reactions (RRTRs) to facilitate convergent peptide assembly and rejuvenate this literature oddity beyond its current obsolete status. We have successfully designed and synthesized heterodipodal peptidosteroid tweezer-type construct, providing a proof-of-concept of our new strategy.

With the sole precedent by DeGrado et al.[11] dealing with the synthesis of linear oligopeptides, we found the possibility of applying the original RRTR methodology on peptide-based conjugates, an emerging area in supramolecular chemistry and multivalency..[12–14] Since the first examples of a three-phase reaction applied in organic chemistry, named as Wolf-and-Lamb and Shadchan chemistry[15–23] and in organometallics[24–27], the methodology of RRTR have lost attention, although it could regain relevance in a contemporary context considering the revived interest in the efficiency of (tandem/one-pot) multi- and interpolymer transformations for high-throughput and flow chemistry[28,29].

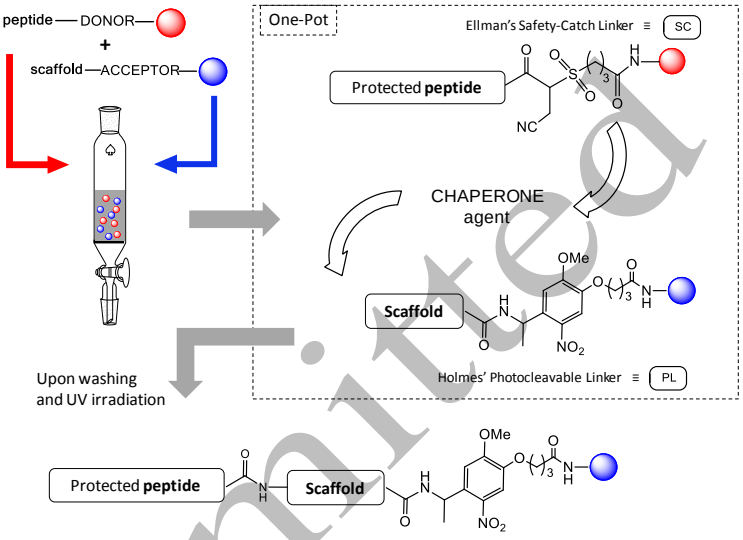


Figure 1. Triphasic Resin-to-Resin Transfer Reaction (RRTR) for the assembly of multivalent peptide conjugates, combining Ellman's Safety-Catch (SC) linker and Holmes' Photocleavable Linker (PL) with a chaperone that mediates solution-phase transfer by acting as both nucleophilic cleavage agent and activated carboxyl leaving group for in situ amidation. Practical simplicity by just mixing resins, collecting the solution which contains the desired conjugate. For monitoring and isolation.....

As presented in Figure 1, resin-to-resin transfer reactions are characterized as triphasic systems involving two solid-phases and a shared solution phase, with transfer between and reaction at both resins as the name implies. In present prototype, we expected a successful match between Ellman's 4-sulfamylbutyryl safety-catch linker[30–32] at the so-called donor resin and Holmes' photolinker moiety[33] at the acceptor resin. Susceptible to nucleophiles upon activation after peptide elongation, the former resin donates a protected segment to an accepting scaffold moiety via the shared solution phase, mediated by a chaperone agent. This chaperone acts as both nucleophile and acyl leaving group with balanced reactivity towards the peptide C-terminus. In situ amidation occurs during interbead exchange by shuttling a (side-chain protected) peptide from the donor resin to a scaffold at acceptor resin in a one-pot triphasic fasion. As such, conjugates are efficiently formed requiring only solid-phase manipulations and no need for prior peptide cleavage needed nor delicate handling of peptide fragments. Excesses of solution-phase species are easily and readily removed after reaction by simple filtration and washing of the resin mixture. Intermediate monitoring and final recovery/isolation is performed likewise by simple filtration upon clean and mild UV irradiation of (sampled) resin[34].

[*] Y. Ruiz García, Dr. D. Verzele, Prof. Dr. A. Madder
Organic and Biomimetic Chemistry Research Group
Department of Organic Chemistry, Ghent University
Krijgslaan 281 (S4), B-9000 Ghent (Belgium)
Fax: (+32) 9 264 4998
E-mail: annemieke.madder@ugent.be

[**] This work was supported by Ghent University, the Fund for Scientific Research (FWO) Flanders, and the Agency for Innovation through Science and Technology (IWT) Flanders. A. M. acknowledges..... Y. R. G. is indebted.....

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

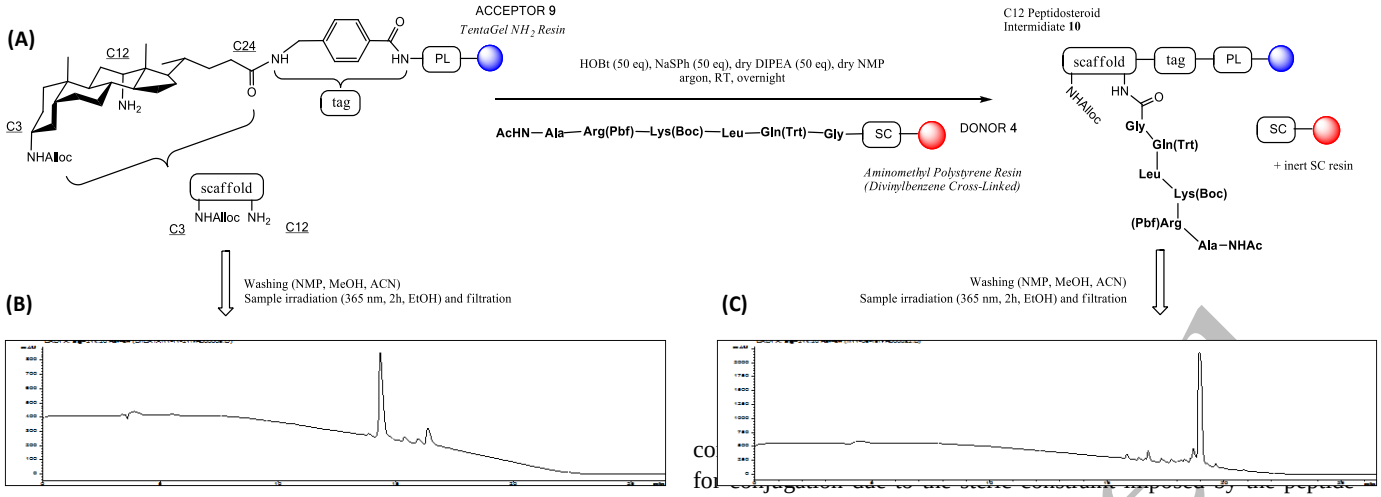


Figure 2. First RRTR to C12 bile acid position. (A) Synthetic conditions. (B) RPHPLC of starting acceptor after photocleavage. (C) RPHPLC after photocleavage showing complete conversion to (side-chain protected) monomeric C12-peptidosteroid conjugate **xx**. (cleavage as C24-carboxyamides, compound number interchangeably, equivalents relative to acceptor resin) (RPHPLC conditions: Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min).

As detailed in Figure 2 (A), our proof-of-concept design comprised of a bile acid based steroid scaffold as accepting template [35] and GCN4 transcription factor derived model peptides shuttled from the donor resin. The employed GCN4 model peptides feature a representative variety of residues and protecting groups. To allow HPLC monitoring of reactions involving essentially UV-inactive intermediates, an aminomethyl benzoic acid moiety featuring a combination of adequate UV-sensitivity and sufficient reactivity (in comparison with its more UV-active *p*-aminobenzoic acid counterpart which is less prone to further amide couplings at the aniline), was installed between the cleavable linker to the solid support and the available C24 position of the bile acid moiety. Figure 2 (B) and (C) illustrate the clean conversion of acceptor resin xx into the C12 peptidosteroid intermediate xx, through reaction with donor resin xx. The Tentagel acceptor resin ensures sufficient swelling in NMP for penetration of the liberated (protected) oligopeptides from the polystyrene donor resin in current prototype. The donor peptide is situated on a preloaded Gly-Safety-Catch (donor) resin and needs to be activated in a suitable fashion for release and transfer onto the acceptor resin. Optimal conditions for activation and release were identified using NaSPh and HOBt as chaperone. Addition of DIPEA was shown to be essential for deprotonation of the accepting amine and thus ensure its nucleophilicity.

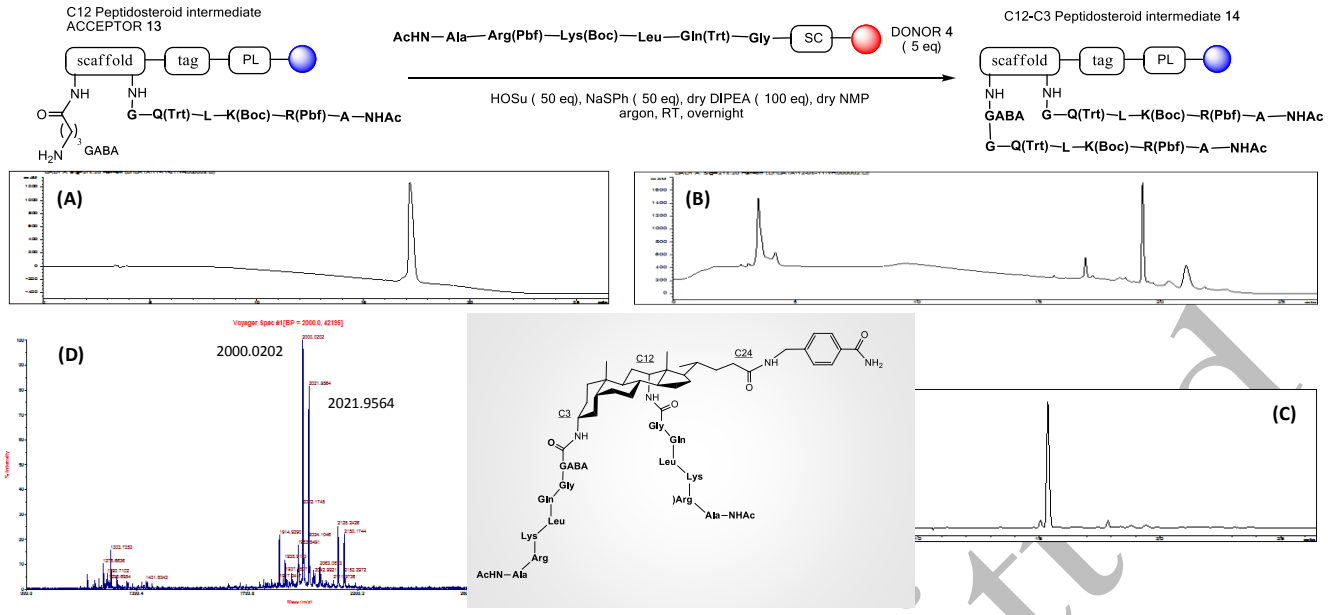
The reasonable excess of donor peptide is easily removed by filtration, which facilitates the purification process in contrast to other convergent, solution-phase schemes. In addition, the final compound is completely isolated from remaining resin excess. The use of a Safety-Catch linker, remaining on the donor resin after RRTR reaction, provides the benefit of inertness during subsequent steps, monitoring and final isolation, and allows standard colour testing for rapid verification next to photolytic cleavage without affecting the final product. Our methodology further offers the possibility to handle minute quantities of precious intermediates as they are immobilized on the resin during the entire reaction time and thus easily manageable without the need of cleavage for further steps.

Deoxycholid acid derived orthogonally protected amine was attached to a solid support, already containing a photolabile linker and a tag. Subsequent conjugation to the deprotected amine at C12 with a protected hexapeptide on solid support delivered the first proof-of-concept of our methodology. Then, Alloc deprotection of amine at the C3 position was carried out with our optimized

attached at C12 position. We observed that the attachment of a spacer was necessary prior RRTR to C3. Consequently, a GABA moiety was installed prior to RRTR to alleviate steric constraints.

In the context of our research towards conformationally defined multipodal peptide architectures based on bile acid derivatives as anchoring scaffolds, both for receptor/loop design and for zipper-type transcription factor models, we have considered the newly developed convergent strategy for the construction of library formats in these areas, accelerating the trial-and-error optimization of binding[36–40]. The steroid scaffold is considered a demanding/challenging building block for decoration due to sterical reasons, as such underscoring the robustness of the optimized conditions. We here present the synthesis of steroid-peptide based conjugates via RRTR. We tested our methodology on a truncated arbitrary model derived from the GCN4 leucine zipper homodimer[41].

For the optimization of conjugation at C3, a variety of chaperone agents were tested for this more challenging position: HOBt, HOAt, HOSu and Oxyma Pure (ethyl 2-cyano-2-(hydroxyimino) acetate)[42]. Again in line with DeGrado precedent, either HOBt or HOSu performed equally well, plus HOSu was a good choice considering commercial scarcity of explosive benzotriazole-based reagents. A similar excess of donor xx was reasonable considering the resin accumulation during the reaction, which can inhibit the cleavage of the peptide from the resin and the posterior conjugation to the acceptor resin. A larger excess of DIPEA was needed, due to the sterically hindered position of the amine at the C3 and the resin accumulation. Monitoring of the intermediate compounds was carried out by RPHPLC. In addition, protected peptidosteroid conjugates could be successfully analyzed on a Jupiter C4 300Å column for polarity reasons of such hydrophobic compounds. After side-chain deprotection, desired final compound could be analyzed in good integrity, with both RPHPLC and MALDI.



As demonstrated in Figure 3. As such RRTR synthesis of dimeric 13-mer, a length often starting to pose difficulties as such during SPPS due to aggregation of peptide chain during synthesis and sterical hindrance imposed by the specific design, can be successfully synthesized with our convergent methodology.

In summary, we here present an upgrade of classical condensation strategy, without the practical encumbrances often associated with synthesizing, isolating, purifying, and handling of the separate peptide fragments to be conjugated. We have achieved the synthesis of a supramolecular, multivalent (bio)conjugate with precise geometrical and appendage composition, through a fully and genuine solid-supported methodology. Moreover, protected peptide fragment manipulations are troublesome due to solubility issues. Our convergent methodology combines the orthogonality of the Ellman's safety-catch linker with the use of Holmes' photolinker to obtain multipolymeric schemes with notable advantages over the last Convergent Solid Phase Peptide Synthesis and can be proposed as an alternative to the former Native Chemical Ligation in peptide assembly in which common schemes and decoration efforts fail. Consequetly, it is proposed as a shortcut towards synthesis of larger and more complex polypeptides and architectures. Furthermore, it compensates for no time-consuming purification.

As main features of our procedure, there are no intermediates in solution as is typical for other ligations and excesses are easily removed. Therefore, handling of intermediate peptidosteroid conjugates is not necessary, avoiding low yields and shortening the synthetic route. The proof-of-concept we present here, is delivered using a scaffold featuring a combination of characteristics that include sterical hindered and concave positions, in a close proximity, that could possibly cause problems such as incomplete couplings with other ligation methodologies. Furthermore, strategic integration of Fmoc, Boc, and Alloc chemistries, combined with nucleophile susceptible Safety-Catch and photolabile cleavage, allowed the obtaining of a practical solid- phase procedure beyond the capabilities of separate approaches.

Figure 3. Second RRTR to C3-position upon Alloc deprotection and GABA introduction, to yield the desired dimeric peptidosteroid upon final deprotection and cleavage. (A) RPHPLC of (side-chain protected) C3-GABA derived acceptor intermediate xx. (B) RPHPLC showing complete C3-RRTR conversion towards side-chain protected C12-C3 dimeric peptidosteroid. (C) RPHPLC and (D) MALDI of desired dimeric peptidosteroid upon final cleavage and side-chain deprotection. (cleavage as C24-carboxyamides, compound number interchangeably, equivalents relative to acceptor resin) (RPHPLC with Jupiter C4 300 Å; gradient 0 to 100% increase of B (versus A) in 15 min and MALDI recorded in the positive and reflection mode)

As mentioned above, advanced orthogonality, simple but fine-tuned, and entirely solid-phase procedure characterized our methodology. Moreover, our design can be adapted for other purposes, as extra decoration position in between the scaffold attachment point and photocleavable linker provides therein an extra dimension for multivalency, to be combined with further RRTR procedures as described herein. Our current research, facing the context of modern (SP)CL and illustrating a contemporary field of multivalent peptide architectures as supramolecular devices, we here wish to disclose our first successes in that direction, with potential broad utility in other designs, easily executed and implemented. In conclusion, we describe the first use of resin-to-resin for multivalent decoration to get conjugates, thereby presenting a proof-of-concept for the untapped opportunities of this methodology towards a wide range of applications in a contemporary context. Our strategy can be useful for solid-phase conjugations amenable to library formats; applicable in areas as diverse as protein models, receptor enzyme design, chemical biology and classical medicinal chemistry, and compatible in microwave and flow chemistry [cross-ref Ley + 2012 papers: Noël Angewandte], consequently, rejuvenating the fashionable interest in such systems.

Experimental Section

Typical procedure for RRTR (illustrated for Cx decoration of xx): text text text text text.

[blijkt niet verplicht en kan volledig in Supporting gezet worden indien gewenst/nodig (character count)]

...as further detailed in the Supporting information.

Received: ((will be filled in by the editorial staff))
Published online on ((will be filled in by the editorial staff))

Keywords: solid-phase synthesis · scaffold decoration · resin-to-resin transfer reactions · peptides · multivalent conjugates

[1] E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 6974–98.

[2] R. K. V Lim, Q. Lin, *Chem. Commun. (Camb).* **2010**, *46*, 1589–600.

[3] J. Y. Lee, D. Bang, *Biopolymers* **2010**, *94*, 441–7.

[4] D. Verzele, A. Madder, *Chembiochem* **2013**, *14*, 1032–48.

[5] H. Ding, A. Shigenaga, K. Sato, K. Morishita, A. Otaka, *Org. Lett.* **2011**, *13*, 5588–91.

[6] P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Science (80-.).* **1994**, *266*, 776–779.

[7] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338–51.

[8] V. R. Pattabiraman, J. W. Bode, *Nature* **2011**, *480*, 471–9.

[9] P. E. Dawson, S. B. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923–60.

[10] F. Mende, O. Seitz, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 1232–40.

[11] Y. Hamuro, M. A. Scialdone, W. F. DeGrado, *J. Am. Chem. Soc.* **1999**, *121*, 1636–1644.

[12] V. Martos, P. Castreño, J. Valero, J. de Mendoza, *Curr. Opin. Chem. Biol.* **2008**, *12*, 698–706.

[13] D. A. Uhlenheuer, K. Petkau, L. Brunsveld, *Combining Supramolecular Chemistry with Biology.*, The Royal Society Of Chemistry, **2010**.

[14] C. H. Battle, J. Jayawickramarajah, *Supramol. Chem. From Mol. to Nanomater.* **2012**, 1885–1908.

[15] H. Kautsky, B. H. de, *Naturwissenschaften* **1931**, *19*, 1043.

[16] J. Rebek Jr., F. Gavina, *J. Am. Chem. Soc.* **1975**, *97*, 3453–3456.

[17] J. Rebek, F. Gavina, *J. Am. Chem. Soc.* **1974**, *96*, 7112–7114.

[18] J. Rebek Jr., *Tetrahedron* **1979**, *35*, 723–731.

[19] J. Rebek Jr., F. Gavina, C. Navarro, *J. Am. Chem. Soc.* **1978**, *100*, 8113–8117.

[20] J. Rebek, D. Brown, S. Zimmerman, *J. Am. Chem. Soc.* **1975**, *97*, 454–455.

[21] J. Rebek, D. Brown, S. Zimmerman, *J. Am. Chem. Soc.* **1975**, *97*, 4407–4408.

[22] J. Rebek, D. Brown, S. Zimmerman, *J. Am. Chem. Soc.* **1975**, *97*, 454–455.

[23] S. Wolf, C. S. Foote, J. Rebek Jr., *J. Am. Chem. Soc.* **1978**, *100*, 7770–7771.

[24] M. Gravel, K. A. Thompson, M. Zak, C. Bérubé, D. G. Hall, *J. Org. Chem.* **2002**, *67*, 3–15.

[25] M. Gravel, C. D. Bérubé, D. G. Hall, *J. Comb. Chem.* **2000**, *2*, 228–231.

[26] J. Tulla-Puche, G. Barany, *Tetrahedron* **2005**, *61*, 2195–2201.

[27] K. A. Thompson, D. G. Hall, *Chem. Commun.* **2000**, 2379–2380.

[28] S. V Ley, I. R. Baxendale, *Nat. Rev. Drug Discov.* **2002**, *1*, 573–86.

[29] S. V. Ley, I. R. Baxendale, *Chim. Int. J. Chem.* **2008**, *62*, 162–168.

[30] B. J. Backes, A. A. Virgilio, J. A. Ellman, *J. Am. Chem. Soc.* **1996**, *118*, 3055–3056.

[31] P. Heidler, A. Link, *Bioorg. Med. Chem.* **2005**, *13*, 585–99.

[32] F. Burlina, C. Morris, R. Behrendt, P. White, J. Offer, *Chem. Commun. (Camb).* **2012**, *48*, 2579–81.

[33] C. P. Holmes, *J. Org. Chem.* **1997**, *62*, 2370–2380.

[34] D. Verzele, A. Madder, *Eur. J. Org. Chem.* **2013**, *2013*, n/a.

[35] D. Verzele, A. Madder, *Eur. J. Org. Chem.* **2007**, 1793–1797.

[36] C. A. Bode, C. P. Muller, A. Madder, *J. Pept. Sci.* **2007**, *13*, 702–708.

[37] A. Madder, L. Li, M. H. De, N. Farcy, H. D. Van, F. Fant, G. Vanhoenacker, P. Sandra, A. P. Davis, C. P. J. De, *J. Comb. Chem.* **2002**, *4*, 552–562.

[38] C. A. Bode, T. Bechet, E. Prodhomme, K. Gheysen, P. Gregoir, J. C. Martins, C. P. Muller, A. Madder, *Org. Biomol. Chem.* **2009**, *7*, 3391–3399.

[39] D. Verzele, S. Figaroli, A. Madder, *Molecules* **n.d.**, *16*, 10168–10186.

[40] M. H. De, A. Madder, N. Farcy, C. P. J. De, M. N. Perez-Payan, L. M. Ohberg, A. P. Davis, *Angew. Chem., Int. Ed.* **2000**, *39*, 145–148.

[41] T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell (Cambridge, Mass.)* **1992**, *71*, 1223–1237.

[42] S. N. Khattab, R. Subirós-Funosas, A. El-Faham, F. Albericio, *ChemistryOpen* **2012**, *1*, 147–52.

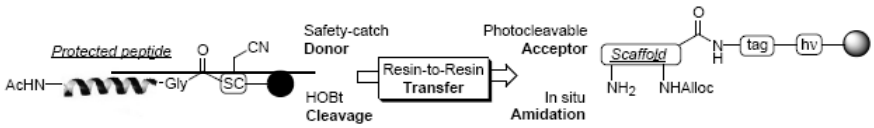
Entry for the Table of Contents

Layout 2:

Resin-to-Resin Conjugation

Yara Ruiz García, Dr. Dieter Verzele and Prof. Dr. Annemieke Madder*

_____ **Page – Page**



Untapped Opportunities of ~~Triphasic~~ Resin-to-Resin Transfer Reactions (RRTR) for the Convergent Assembly of Multivalent Peptide Conjugates

((Text for Table of Contents, max. 450 characters))

[TOC Graphical als eerste voorbeeld: werken met kleurtjes sowieso; HOBt/HOSu?; herwerken naar dimeer ipv huidig; picturaler met minder woordjes]

ChemComm

RSCPublishing

COMMUNICATION

Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake†

Cite this: DOI: 10.1039/xoxxoooooox

Abhishek Iyer[a], Dorien Van Lysebetten[a], Yara Ruiz García[a], Benoit Louage[b], Bruno De Geest[b], and Annemieke Madder*[a]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

The basic DNA recognition region of the GCN4 protein comprising 23 amino acids has been modified to contain two optimally positioned cysteines which have been linked using cross-linkers of suitable lengths. This results in stapled peptides each with a stabilized α-helical conformation and concurrent enhancement of DNA binding and cellular uptake.

Transcription factors are usually classified according to the fold of their DNA-binding domains and grouped into a small number of families, like the bZIP, bHLH, homeodomain, HTH, and zinc fingers which have already been studied in detail[1]. Most transcription factors are large proteins possessing complex secondary structures rendering it difficult to design smaller synthetically accessible versions thereof retaining the DNA binding capacity. The GCN4 leucine zipper is however a relatively easy protein to mimic due to its well-defined dimerization domain and basic recognition region. Construction of smaller dimeric mimics of this transcription factor, although feasible, have proven to be synthetically challenging[2,3].

Previous attempts at more thorough structural minimisation and reduction of complexity using the monomeric GCN4 peptide has shown that DNA binding is greatly reduced due to loss of secondary structure[4]. Indeed, since the basic region of the GCN4 transcription factor cannot adopt a helical fold in solution by itself nor bind to DNA[5] due to entropic reasons,[6] an external factor for forcing the peptide into a helical conformation is needed. It has already been shown that dimerization can enhance DNA binding. For this reason, unlike the existing synthetic bZip peptides in which DNA binding is induced via dimerisation,[7, 2] we here aim to stabilize a single α helix via peptide stapling, reasoning that enhancing the helicity within the monomer should sufficiently stabilize the conformation to allow DNA binding[4].

Stapled peptides have been used extensively for improving helicity[8], increasing cell-penetration[9,10], proteolytic stability and enhancing peptide-protein interactions (PPIs)[11]. In all the above mentioned cases stapled peptides have proven to be an effective methodology[12]. However, the use of stapled peptides for enhancing DNA binding has remained largely unexplored. In this work we have examined the DNA binding induced by stapling using the monomeric GCN4 transcription factor as a model peptide, by comparing the i, i+4 and i, i+7 stapling methods and varying the positions of the staple along the helix. The cellular uptake of the constructs was also investigated using fluorescently labelled versions of the peptide.

Initial studies were dedicated to selecting the most suitable method for peptide stapling among the ones reported in literature. The idea was to adopt a method as general as possible to be potentially applicable to any DNA binding peptide to increase its helicity. To ensure easy and cost-effective modification the use of unnatural amino acids was avoided. In case of the very well-known hydrocarbon stapling, coupling of the non-natural amino acids as well as peptide folding on resin can be a problem[13]. The commercial availability and/or easy synthesis of cross-linking moieties, mild reactions conditions and easy scalability due to synthesis in solution, were the reasons for choosing cysteine cross-linking for current study[12].

A detailed analysis of the essential contacts for DNA binding as derived from the reported crystal structure was described earlier by Ellenberger and Keller.[14, 15] Based on this study, amino acids indicated as not involved in DNA contacts were identified and systematically replaced by Cys according to an i,i+4 or i,i+7 format. Molecular modelling aided visualisation based on the pdb file 1YSA was used to ensure that the introduced linkers point away from the DNA and not towards it thereby avoiding any steric repulsion which may arise due to peptide stapling. In this way three different peptides, comprising

[a] *Organic and Biomimetic Chemistry Research Group, Krijgslaan 281, S4, B-9000 Gent, Belgium. E-mail: annemieke.madder@ugent.be; Fax:+32-9-264 49 98*
[b] *Faculty of Pharmaceutical Sciences. Department of Pharmaceutics. Ghent University. Harelbekestraat 72, 9000 Ghent. Belgium. Tel.: +32 9 264 80 55. E-mail: br.degeest@ugent.be*

† Electronic supplementary information (ESI) available: Experimental procedures, HPLC and NMR data. See DOI: 10.1039.xxxx

the D_{226} – Q_{248} sequence from the DNA binding basic region of GCN4, containing a double Cys substitution (at positions 237/244 for **1**, 229/233 for **2** and 233/237 for **3**) were synthesized on solid support, cleaved and subsequently treated with various linkers yielding a series of five stapled peptides as shown in Figure 1.

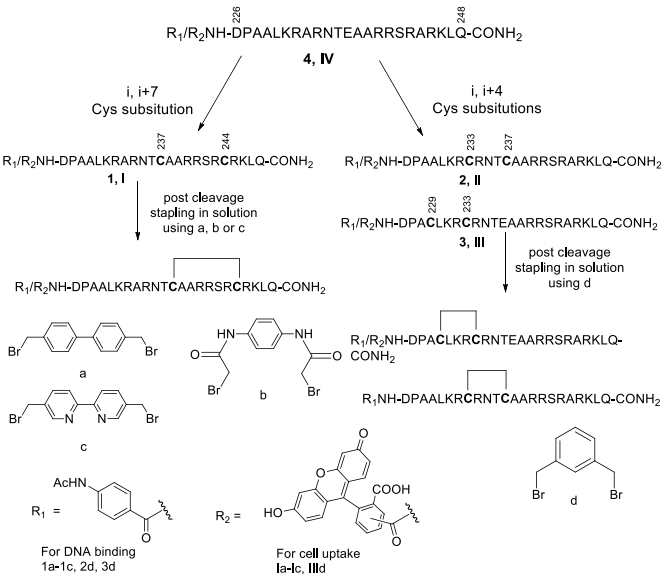


Fig.1: Synthetic stapled peptides **1a-c, 2d, 3d** for DNA binding and unmodified basic region peptide **4**. The peptides are N-terminally capped with a p-acetylamino benzoic acid (ABA) moiety to ensure UV-based detection and analysis in the case of DNA binding. For cell uptake studies the ABA moiety was replaced by fluorescein.

The stapled peptides and the unmodified WT basic region peptide were successfully synthesized. The cross-linkers **a, c** & **d** are commercially available. The stapling moiety **b** has not been used thus far for peptide stapling and was designed and synthesized as a more polar alternative to the biphenyl and bipyridine cross-linkers.

Next, the DNA binding capacity of all peptides was evaluated through EMSA titration of various peptide concentrations to the cognate CRE DNA sequence 5' – CGG ATG ACG TCA TTT TTT TTC – 3' (Figure 2).

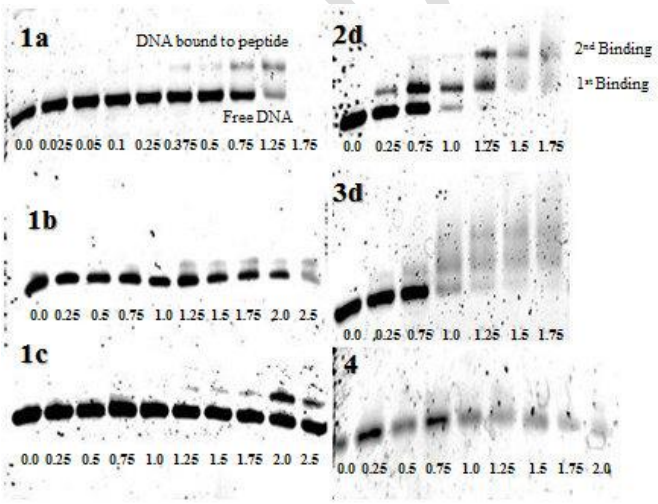


Fig. 2: EMSA titrations for peptides **1a-c, 2d** & **3d**. Loading mixture comprises 5 µL from a mixture of 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide resulting in a total DNA concentration of 167 nM. The loading buffer consists of 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M. Peptide concentrations from left to right (in µM) are indicated below each gel.

The WT peptide **4** does not bind DNA under the given EMSA conditions. For all the synthetic constructs **1a-c, 2d** & **3d** we see enhanced DNA binding as compared to peptide **4**. In the case of peptides **1a-c** binding is only observed at higher concentrations of peptide as compared to **2d** and **3d**. The more hydrophobic nature of the biphenyl cross-linker, makes peptide **1a** more susceptible to aggregation as can be seen from the complete disappearance of the bands in the last lane. On the other hand, peptides **1b** and **1c** are less prone to aggregation but do not bind as well as **1a**. Unlike peptides **1a-c**, the DNA binding pattern in gel **2d** is different. We propose that the appearance of two bands related to DNA-peptide complex formation is due to the presence of two binding sites in the CRE sequence (…GTCAT…). It can be noticed that after full occupation of one binding site by the peptide, remaining non-bound peptide can interact with the second binding site. The binding pattern of peptide **4** is unique and only visible in the case of the two sequences that bind at a lower concentration range. It can further be noticed that peptide **5** shows a similar binding pattern as peptide **4** but suffers from non-specific interactions to a small extent. Non-specific interactions cause aggregation, bands getting blurred or disappearance of all bands.

From the few examples about DNA binding stapled peptides that do exist in literature[4,16] and from our own data we believe the major challenge in case of DNA binding peptides is that increasing the helicity has to be complemented with a degree of flexibility in order to account for the conformational change which occurs when the peptide binds to DNA. Since this conformational change is more significant in the case of DNA binding peptides as compared to PPIs, merely locking the conformation of the peptide into a helix may not be sufficient to achieve enhanced DNA binding. On the other hand, peptide stapling by providing an N or C terminal helix stabilization can give better results in the case of DNA binding peptides as can be seen from the binding pattern of peptides **2d** & **3d**. This, of course, may result in a less helical conformation of the peptide without the presence of DNA but allows it to adjust its conformation into a structure whereby it can maximize the contact between the positively charged side chain interactions mainly involving the Lys and Arg residues with the negatively charged backbone of the DNA without a high entropic penalty. We believe peptide **2d** fits this criteria and hence is the best DNA binder from the constructs synthesized. Moreover, it is able to bind in a dimeric fashion without having been artificially dimerized. We thus observe for the first time dimeric DNA binding with a monomeric GCN4 derived basic region peptide.

It has been postulated that stabilization of the secondary structure via peptide stapling can enhance cell uptake[17]. Peptides with hydrocarbon staples in particular have shown considerable increase in uptake as compared to their non-stapled counterparts[10,18]. The fluorescently labelled peptides **Ia-c** & **IId** were tested in a cellular environment using RAW 264.7 mouse macrophages. Confocal

microscopy confirmed that cell penetration is achieved at a concentration as low as 0.25 µM for all peptides at 37°C, including native GCN4 sequence **IV**, as seen from figure 3A-E. The cell penetrating properties even at low concentrations combined with the low cytotoxicity values observed in the MTT-assay (see SI) make these a valuable set of CPPs.

Incubation at 4 °C provides insight into the mode of uptake as endocytic pathways are shut down at this temperature. Comparison of the data obtained at 4 °C and 37 °C with flow cytometry (figure 5) confocal microscopy (see SI) showed that active uptake is the main internalization pathway as the uptake drops at 4 °C.

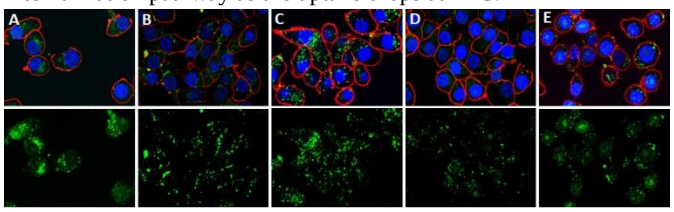


Fig. 3: Confocal microscopy images of the uptake of peptides (A) **Ia**, (B) **Ib**, (C) **Ic**, (D) **IIId**, and (E) **IV** at 37 °C. The upper panel shows accumulated images of DNA (blue), cell membrane (red) and fluorescein (green). The lower panel shows the fluorescein image. (Scale bar = 20 µm).

Quantification of the uptake by flow cytometry shows a concentration dependent uptake (figure 5). The mean fluorescence emitted by the fluorescent labelled peptides in the cell increased considerably for an i, i+7 staple.

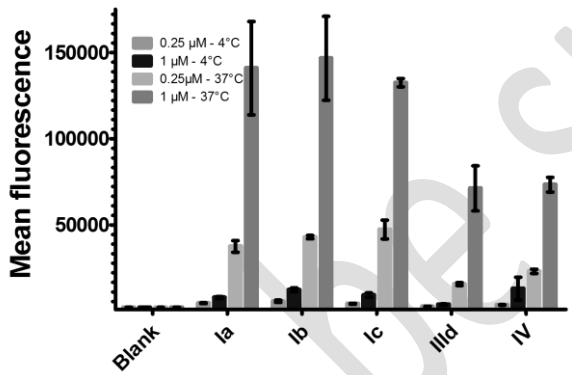


Fig. 4: Mean fluorescence of fluorescently labelled peptides by incubation at 0.25 µM.

In conclusion, the elegant nature of the synthesis in combination with the observed DNA binding and cellular uptake properties render these constructs of considerable and specific interest among the mimics of the GCN4 transcription factor reported to date. Through this work and due to the nature of the stapled peptides, we believe that a general method is now at our disposal to allow DNA binding and enhance cellular uptake of a given DNA binding peptide while avoiding tedious synthetic routes. It has been further shown that N-terminal helix stabilization as in the case of **2d** is more effective for enhancing DNA binding than stapling in the middle of the sequence or towards the C-terminus (**1a-c**). For cell penetration, however, the more helical i, i+7 stapled peptides **Ia-c** have shown better uptake than **IV**. Although the two are not mutually exclusive, for future applications such as DNA binding in cellulo a balance will have to exist between a peptide's DNA binding and cell penetration abilities.

Acknowledgements

Tim Courtin is acknowledged for the determination of the concentration of the final compounds by ERECTIC-NMR. Jan Goeman is acknowledged for the LC-MS analysis. Abhishek Iyer and Yara Ruiz García are indebted to the Marie Curie Early Stage Research Training Fellowship of the European Community's Seventh Framework Programme under contract number PITN-GA-2010-238679. The FWO and BOF are also acknowledged for financial support.

Notes and references

1. E. Pazos, J. Mosquera, M. E. Vµzquez, and J. L. Mascareças, *Chembiochem*, 2011, **12**, 1958–1973.
2. T. Morii, M. Simomura, and S. Morimoto, *J. Am. Chem. Soc.*, 1993, **32**, 1150–1151.
3. A. Jimønez-balsa, E. Pazos, B. Martínez-albardonedo, J. L. Mascareças, and M. E. Vµzquez, 2012, 1–6.
4. M. I. N. Zhang, B. Wu, H. Zhao, and J. W. Taylor, *J. Pept. Sci.*, 2002, **136**, 125–136.
5. C. Park, J. L. Campbell, and W. A. Goddard, 1996, **7863**, 4892–4896.
6. L. K. Henchey, A. L. Jochim, and P. S. Arora, *Curr. Opin. Chem. Biol.*, 2008, **12**, 22–26.
7. J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vázquez, and J. L. Mascareñas, *Nat. Commun.*, 2013, **4**, 1874.
8. K. Estieu-gionnet and G. Guichard, *Drug Discov.*, 2011, **6**, 937–963.
9. M. M. Madden, A. Muppidi, Z. Li, X. Li, J. Chen, and Q. Lin, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1472–1475.
10. H.- Inhibitor, H. Zhang, Q. Zhao, S. Bhattacharya, A. A. Waheed, X. Tong, A. Hong, S. Heck, F. Curreli, M. Goger, D. Cowburn, E. O. Freed, and A. K. Debnath, *J. Mol. Biol.*, 2008, **378**, 565–580.
11. C. J. Brown, S. T. Quah, J. Jong, A. M. Goh, P. C. Chiam, K. H. Khoo, M. L. Choong, M. a Lee, L. Yurlova, K. Zolghadr, T. L. Joseph, C. S. Verma, and D. P. Lane, *ACS Chem. Biol.*, 2013, **8**, 506–12.
12. H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. DeGrado, and D. C. Greenbaum, *J. Am. Chem. Soc.*, 2012, **134**, 17704–13.
13. L. D. Walensky and G. H. Bird, *J. Med. Chem.*, 2014, **57**, 6275–6288.
14. T. E. Ellenberger, C. J. Brandl, K. Struhl, and S. C. Harrison, 1992, **71**, 1223–1237.
15. W. Keller, P. Ko, and T. J. Richmond, 1995, 657–667.
16. L. Guerrero, O. S. Smart, G. A. Woolley, and R. K. Allemann, *J. Am. Chem. Soc.*, 2005, **127**, 15624–15629.
17. A. Muppidi, Z. Wang, X. Li, J. Chen, and Q. Lin, *Chem. Commun. (Camb).*, 2011, **47**, 9396–8.
18. W. Nomura, H. Aikawa, N. Ohashi, E. Urano, M. Métifiot, M. Fujino, K. Maddali, T. Ozaki, A. Nozue, T. Narumi, C. Hashimoto, T. Tanaka, Y. Pommier, N. Yamamoto, J. a Komano, T. Murakami, and H. Tamamura, *ACS Chem. Biol.*, 2013, **8**, 2235–44.

Annex III:

Curriculum Vitae

Yara Ruiz

■ Profile

Full name Yara Ruiz García
Date of birth 16/10/1987
Marital status Single
Nationality Spanish
Address Vossenweg 25
1820 Steenokkerzeel (Belgium)
Mob. +32491242339
E-mail yruizg7@gmail.com

Personal aim: I would like to have the opportunity to apply my personal skills as an experienced researcher in organic chemistry.

Key words: Supramolecular Chemistry Organic Chemistry Peptide Chemistry

■ Working experience

Jan. 2015 Process development and scale-up. Peptisyntha SA - Cordenpharma (Brussels).

April-Sep. 2014 Research Assistant PostDoc.
School of Life Sciences. University of Lincoln (UK). Promoter: Prof. Ishwar Singh.
Drug development processes. A novel anchoring approach as platform technology for design and improved delivery for a new class of broad spectrum antibacterials.

■ Education

2010 - Now PhD Student at the Department of Organic and Macromolecular Chemistry. Ghent University (Belgium)
Theme: Phosphate Transfer and Recognition: Towards DNA binding and cleaving construct mimics.
PhosChemRec Network. Marie Curie Early Stage Research Training Fellowship.
Promoter: Prof. Dr. Annemieke Madder.

Projects:
DNA recognition by artificial DNA binders
DNA cleavage
Peptide-based hydrogels for controlled-released technology and pain management
Design and synthesis of artificial receptors for detection of apoptosis
Resin-to-resin transfer reactions for the convergent assembly of multivalent peptide-conjugates
Development of a synthetic receptor for trace determination of Endocrine Disruptor Chemicals

2010 - 2013 Doctoral School of Natural Science (Ghent University). 60 ECTS.

2005 - 2010 Master of Chemistry (Organic Chemistry and Physical Chemistry)
Universidad de Valladolid (Spain)

2005 Secondary education (graduate with special distinction)
Colegio La Salle Palencia (Spain)

■ Professional Experience

Solid Phase Organic Synthesis	Toxin receptors	Molecular modelling
Solid Phase Peptide Synthesis	Halogen and anion receptors	Pharmaceutical bioinformatics
Synthesis of peptide conjugates	Microwave synthesis	Click chemistry
Organic synthesis	Peptide Hydrogels	DNA recognition and cleavage

Analytical techniques: MS(ESI), MALDI-TOF, HPLC, LC-MS(ESI), EMSA, circular dichroism

■ Related experience

2009 Summer Internship in the Physics and Chemistry of Nanostructures Research Group. Ghent University (Ghent - Belgium).

2008 Job Student in Becton & Dickinson (Erembodegem – Belgium), working on "The REACH Project".

2008 Summer Internship Grant in SEDA Solubles S.L. (Palencia-Spain). Quality Department.

2008 Laboratory assistant in the organic chemistry department at the University.

2003-2005 Laboratory assistant and Chemistry tutor in Colegio La Salle Palencia.

■ Skills

Motivation	Patience	Ambition
Creativity	Hard-worker	Out-going
Team-working	Organization	Communicative
Responsible		

■ Publications

H. Diaz-Velazquez, Y. Ruiz-García, M. Vandichel, F. Verpoort and A. Madder. 'Water-Soluble Bis-NHC-Cu (I) Catalysts for the Preparation of Triazoles via CuAAC. Mechanistic Analysis and Bioconjugation Experiments in Aqueous Media'. Organic and Biomolecular Chemistry. Accepted. 2014. DOI: 10.1039/C4OB01350F.

M. Sanders, Y. Guo, Abhishek Iyer, Y. Ruiz García et al. 'An immunogen synthesis strategy for the development of specific anti-deoxynivalenol monoclonal antibodies'. Food Additives and Contaminants: part A. Accepted. 2014. DOI: 10.1080/19440049.2014.955887.

Y. Ruiz Garcia, D. Verzele and A. Madder. 'Convergent approaches on solid phase synthesis: Towards biomimetic transcription factors'. Journal of Peptide Science. (2012). 18, S1, 117-118.

Y. Ruiz García, V. Pabón, E. Smith and A. Madder. 'Sequence-Specific Recognition of a Nonpalindromic DNA Sequence by a Heterodimeric Transcription Factor Mimic'. To be submitted.

Y. Ruiz García, V. Pabón, E. Smith, T. Kraus and A. Madder. 'Sequence-Specific DNA Recognition by bZIP-Leucine Zipper Transcription Factor Models with Different Dimerization Interface'. To be submitted.

Y. Ruiz-García, Abhishek Iyer, Edvard Smith and A. Madder. 'Sequence-Selective DNA Recognition by Unprecedented Peptide-Bile Acid Conjugates via CuAAC'. To be submitted.

Y. Ruiz-García, D. Verzele and A. Madder. 'Untapped Opportunities of Resin-To-Resin Transfer Reactions for the Convergent Assembly of Multivalent Peptide Conjugates'. To be submitted.

■ Presentations

Y. Ruiz García, Annemieke Madder. "Design and Synthesis of Artificial DNA Binders: Towards Sequence-Specific DNA Recognition"

NAME OF CONFERENCE	LOCATION	DATE
Belgian Peptide Group Meeting	**Ghent (BE)**	February 10th-11th (2014)

Y. Ruiz García, Annemieke Madder. "Design and Synthesis of Artificial DNA Binders: Towards Sequence-Specific DNA Recognition"
Sigma-Aldrich Symposium **Blankenberge (BE)** December 5th-6th (2013)

Y. Ruiz García, Annemieke Madder. 'Recognition of Phosphates: From DNA recognition to artificial DNA nucleases'
6th Network Meeting PhosChemRec **Cambridge (UK)** August-September 29th-2nd (2013)

Y. Ruiz García, Annemieke Madder. 'Recognition of Phosphates: From DNA to phosphatidylserine'
5th Network Meeting PhosChemRec **Frankfurt (GE)** January 15th – 18th (2013)

Y. Ruiz García, A. Madder. "Convergent Approaches towards cholic acid-based receptors for the recognition of large and small phosphates"
Midterm PhosChemRec Meeting **Ghent (BE)** May 23rd, 2012

Y. Ruiz García, A. Madder. "PATA linker as a moiety facilitating the synthesis of transcription factor mimics via click chemistry"
3rd PhosChemRec Network Meeting **Stockholm (SW)** July 3rd, 2011

Y. Ruiz García, A. Madder. "MW-Assisted 'Click' Synthesis of Cholic Acid-Based Artificial Receptors and Transcription Factor Mimics"
2nd PhosChemRec Network Meeting **Padova (IT)** February 18th, 2011

■ Teaching Experience

Supervisor of Master Thesis : Peptide-based hydrogels for controlled-released technology and pain management. September 2013-June 2014

Supervisor of Master Thesis : Molecular simulation and synthesis of artificial receptors for beauvericine. September 2012-May 2013

Supervisor of Master Thesis: Synthesis of receptors for molecular recognition of beauvericine and cereluide toxins. February-June 2012

Internship supervisor. « MycoHunt project » May- August 2011

■ Awards

2014 Symposium travel grant to attend the 33rd European Peptide Symposium
Research project selected for oral presentation at the 33rd European Peptide Symposium
Research project selected for oral presentation at the 2nd Belgian Peptide Group Meeting

2013 Research project selected for oral presentation at the 17th Sigma-Aldrich Symposium.
Grant to participate in the European Winter School in Physical-Organic Chemistry. 27th-1st February. Bressanone (Italy)

2010 Early Stage Research Grant**.** European Community's Seventh Framework Programme; PITN-GA-2010; 238679. (November 2010-February 2014).
2009 Erasmus Training Grant. July-September.
2005 Spanish Ministry of Education. Grant: First year enrollment at the University.
2005 Greatest Distinction. Secondary Education.

■ Poster Presentations

Yara Ruiz García and Annemieke Madder.
Recognition Of Large And Small Phosphates: Artificial Receptors Through Convergent Ligations. Belgian-German Macromolecular Meeting. 3rd-4th December 2012. Houffalize, Belgium.

Yara Ruiz García, Dieter Verzele and Annemieke Madder.
Development of an Unprecedented Resin-to-Resin Transfer Reaction for Application to Peptide Conjugate Systems. **32nd European Peptide Symposium**. September 2nd-7th 2012. Athens, Greece

Yara Ruiz García, Dieter Verzele, Lieselot Carrette, Jos van den Begin and Annemieke Madder.
Convergent Approach On Solid Phase: Towards Transcription Factor Mimics. ChemCYS conference. March 1st-2nd 2012. Blankenberge, Belgium

Yara Ruiz García, Dieter Verzele and Annemieke Madder.
Towards Transcription Factor Mimics. Sigma-Aldrich Organic Synthesis Meeting. December 1st-2nd 2011. Spa, Belgium.

Yara Ruiz García, Dieter Verzele, Lieselot Carrette, Jos van den Begin and Annemieke Madder.
Convergent Approach On Solid Phase: Towards Transcription Factor Mimics. Annual One-Day Meeting on Medicinal Chemistry of SRC & KVCV (Medchem 2011). November 25th 2011. Gent, Belgium

Yara Ruiz García, Lieselot Carrette, and Annemieke Madder.
Organic and Biomimetic Chemistry of Phosphate Transfer and Recognition. **Marie Curie Researchers Symposium**. September 25th-28th 2011. Warsaw, Poland

■ Non-related experience

2006-2008 Voluntary work with physical and mentally disabled people in Centro Villa San José Palencia.
2006 Voluntary work with physical and mentally disabled people in C.A.M.P, Hermanas Hospitalarias, Burgos, during the summer.
2005-2006 Voluntary work in "Cáritas Diocesana Palencia" helping children with learning problems.
2005 Young workcamp in Asturias (Spain).

■ Hobbies

Latin dances
Contemporary history
Travel
Chemistry

■ Language skills

Spanish	Mother tongue
English	Proficiency level. Certificate in Advanced English (2009) (Cambridge University)
Dutch	Intermediate level. B2 at the Language Center of Ghent University (2012)
French	Intermediate level

■ References

Prof. Dr. Annemieke Madder
Ghent University
Department of Organic and Macromolecular Chemistry
Organic and Biomimetic Chemistry Research Group
Krijgslaan 281 S4
9000 Ghent (Belgium)
Tel: +32 9 264 44 72
annemieke.madder@ugent.be

Dr. Ishwar Singh
University of Lincoln
School of Pharmacy
College of Science
Brayford Pool
LN6 7TS, Lincoln (United Kingdom)
Tel: +44 1522 88 6915
isingh@lincoln.ac.uk